Graduate Texts in Mathematics

GTM

Rabi Bhattacharya
Edward C. Waymire

Continuous Parameter Markov Processes and Stochastic Differential Equations

Springer

Graduate Texts in Mathematics 299

Graduate Texts in Mathematics

Graduate Texts in Mathematics bridge the gap between passive study and creative understanding, offering graduate-level introductions to advanced topics in mathematics. The volumes are carefully written as teaching aids and highlight characteristic features of the theory. Although these books are frequently used as textbooks in graduate courses, they are also suitable for individual study.

Rabi Bhattacharya • Edward C. Waymire

Continuous Parameter Markov Processes and Stochastic Differential Equations

Springer

Rabi Bhattacharya
Department of Mathematics
University of Arizona
Tucson, AZ, USA

Edward C. Waymire
Department of Mathematics
Oregon State University
Corvallis, OR, USA

ISSN 0072-5285 ISSN 2197-5612 (electronic)
Graduate Texts in Mathematics
ISBN 978-3-031-33294-4 ISBN 978-3-031-33296-8 (eBook)
https://doi.org/10.1007/978-3-031-33296-8

Mathematics Subject Classification: 47D07, 60H05, 60H30, 60J25, 60J27, 60J28, 60J35, 60J70, 60J74, 60J76, 60E07, 60F17, 60G44, 60G51, 60G52, 60G53

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Dedicated to Gouri (In loving memory), to Linda, and to Vijay K Gupta (In memory of a dedicated scientific collaborator and dear friend to both of us.)

Preface

The subject of Markov processes in continuous time has an elegant and profound mathematical theory and a great diversity of applications in all fields of natural and social sciences, technology, and engineering. It is our objective to present this theory and many of its applications as simply as we can. Apart from a graduate course in analysis and one in probability, the book is essentially self-contained. For the probability background, A Basic Course in Probability Theory (2016), referred to as BCPT throughout the text, suffices. A few technical results from analysis are proved in Appendices. There are rare but occasional references to the companion stochastic process texts (Bhattacharya and Waymire 2021, 2022). Alternative references can be used as well.

The first chapter reviews Doob's martingale theory as an essential tool used throughout the book. We begin then with an exposition of how a semigroup of transition operators T_t $(t > 0)$, describing the expected evolution in time t of all appropriate functions $f(x)$ of an initial state x, is specified by the derivative of $T_t f(x)$ at time $t = 0$. This derivative A, called the infinitesimal generator, embodies the law of evolution determining the Markov process. The generator A is in integro-differential form, including the class of all second order linear elliptic operators, indicating a deep connection with partial differential equations.

Next, after establishing a broad criterion for right continuity of sample paths, two classes of Markov processes are presented both analytically, using semigroups, and probabilistically, describing pathwise behavior. Jump Markov processes are those that remain in a state for an exponentially distributed random time before moving to another state. Parameters of the exponential times depend on the holding state. Poisson and compound Poisson processes, as well as continuous time

birth-and-death chains, are examples. This is followed by a chapter on processes with independent increments, also called Lévy processes. This class includes the ubiquitous Brownian motion. Leaving that out, the fascinating Lévy-Khinchin-Itô theory exhibits the process as an accumulation of jumps from a Poisson random field.

The most important continuous time stochastic processes are diffusions-which are real or vector-valued Markov processes having continuous sample paths. Much of the book is devoted to this topic. A diffusion may be thought of as locally Brownian, and Itô's beautiful and rich theory of stochastic integrals and stochastic differential equations for the construction of general diffusions is developed around this theme. A centerpiece of this theory is Itô's Lemma, with stochastic differentials and integrals replacing Newtonian calculus. Fairly complete asymptotic properties of these processes in one dimension, due to Feller, such as transience, null and positive recurrence, explosion, are presented in several chapters.

In view of its general importance in Markov process theory, a rather detailed account is presented of the speed of convergence to equilibrium of Harris recurrent Markov processes in discrete time, and one-and-multi-dimensional diffusions in continuous time, using the method of coupling.

Multi-dimensional extensions of Feller's criteria, announced by Khaśminskii, are derived in two chapters under broad conditions on the coefficients of the elliptic infinitesimal generator. Also developed in this text is a widely applicable functional central limit theorem (FCLT) for general ergodic Markov processes. It says, in essence, that the FCLT holds for integrals of all square integrable functions in the range of the infinitesimal generator. This provides in many instances a computable dispersion matrix for the asymptotic Brownian motion. Among several applications, we give a probabilistic derivation of the Taylor-Aris theory of solute transport in a capillary, and the asymptotic Gaussian distribution of a scaled diffusion on $\mathbb{R}^k$ with periodic coefficients.

A transformation rule, known as the Cameron-Martin-Girsanov Theorem, computes the Radon-Nikodym derivative of the distribution on finite intervals $[0, T]$ of one diffusion with respect to another with the same dispersion coefficient but different drifts. This has many uses, one being the proof that a diffusion with non-singular diffusion matrix has full support.

In one dimension, the Hille-Yosida Theorem for semigroups is the cornerstone of Feller's theory for the construction of all one-dimensional regular diffusions. Feller's seminal theory also characterizes all possible boundary conditions as well. An exposition of this theory in a graduate text is one of the highlights of this book. The probabilistic understanding of the main boundary conditions is also emphasized.

Our approach to the construction of certain classes of Markov processes as functions of Markov processes, yields, among others, the construction of reflecting and periodic diffusions and diffusions on a torus. A chapter on Paul Lévy's local time of Brownian motion is included, leading to applications such as his construction of reflecting Brownian motion, as well as the intriguing behavior under mixed boundary conditions. The treatment of local times uses some material from a chapter on integration with respect to L^2-martingales that precedes it. This latter chapter contains the Doob-Meyer decomposition as well.

The intimate connection between stochastic differential equations and elliptic and parabolic partial differential equations of second order, presented in two chapters, attests to the mathematical depth of this part of probability and is used extensively for the asymptotic theory that follows.

The chapters on special topics include one on multi-phase homogenization of periodic diffusions with a large spatial parameter "a." Two examples illustrate this fascinating theory. One displays the passage of the solute profile through Gaussian and non-Gaussian phases as dispersivity rapidly increases with time, while the other has different Gaussian phases, but no substantial increase in dispersion with time. This topic is motivated by widely observed growth in dispersion of solutes in heterogeneous media such as aquifers. Another special topics chapter titled "Skew Brownian Motion" is also motivated by solute transport in porous media. Here also, interesting anomalous asymptotic behavior occurs when the diffusion has change points in the dispersion parameter.

Another special topics chapter is devoted to a brief exposition of the Stroock-Varadhan martingale problem, with an application to the approximation of a one-dimensional diffusion by a scaled birth-and-death chain.

The authors are indebted to Springer Editors Loretta Bartolini and Elizabeth Loew for their guidance, and also to anonymous reviewers for their valuable suggestions. We would like to thank our colleagues Sunder Sethuraman, William Faris, and Enrique Thomann for their continued advice and encouragement. We also thank the University of Arizona graduate students Duncan Bennett and Eric Roon for technical assistance with some TIKZ and Latex code. The authors gratefully acknowledge partial support (DMS1811317, DMS-1408947) from the National Science Foundation during the preparation of this book.

Tucson, AZ, USA Rabi Bhattacharya
Corvallis, OR, USA Edward C. Waymire
July 17, 2023

Course Suggestions

(A) Discontinuous Markov Processes: 1-4, 5, 26

(B) Stochastic Differential Equations: 1, 6-9,10,11,12, 20, 23

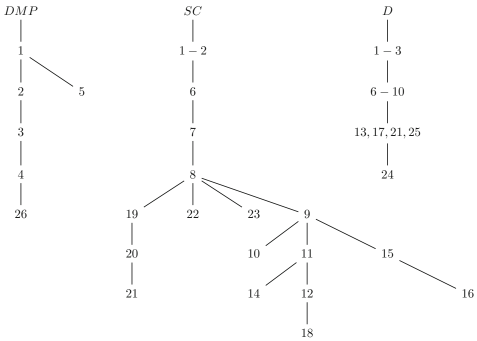
(C) Diffusions and PDEs: 2, 8,9,13, 15-16, 21, 22, 25

(D) Equilibria and Speed of Convergence: 11, 14,18, 22

(E) Fluctuation Laws and Approximation in Distribution: 1,17, 24, 25

Chapter Dependency Diagram

DMP = Discontinuous Markov Processes
SC = Stochastic Calculus
D = Diffusion



Chapter Relationships and Dependencies

A Trilogy on Stochastic Processes

- *Random Walk, Brownian Motion and Martingales* (GTM 292)
- *Stationary Processes and Discrete Time Markov Processes* (GTM 293)
- *Continuous Time Markov Processes and Stochastic Differential Equations* (GTM 299)

The trilogy provides a systematic exposition of stochastic processes—a branch of mathematics that increasingly impacts every area of science and technology.

GTM 292 This volume begins with a broad coverage of random walk, Brownian motion, and martingales—three elegant pillars of probability theory. Brownian motion with its universality shines as a star in the center of probability, with the scaled random walk and the widely applicable martingale functional central limit theorem (FCLT) converging to it. There is an extensive treatment of branching processes with many applications, along with a branching-like treatment initiated by Le Jan and Snitzman of the most basic equations of fluid dynamics, namely, Navier-Stokes equations; it is among the most challenging unsolved problems of mathematics with broad implications to physics. Discrete renewal theory, using coupling, lays the foundation for analyzing the speed of convergence to equilibrium for general Markov processes—a recurring theme in the trilogy. Blackwell's general renewal theory is essential for the computation of ruin probabilities in insurance, especially for heavy-tailed claim sizes which may aggravate the chance of a collapse.

GTM 293 The first part of the second volume is a detailed treatment of Kolmogorov's prediction theory for processes with only weak stationarity, namely, that of second order moments. Together with Wiener's independent result for such processes in continuous time, this has been one of the most important developments in the engineering literature. This is followed by Birkhoff's ergodic theorem for strictly stationary processes which, historically, had provided the affirmation of the long anticipated equality of the time average and the phase average of systems under ergodicity in statistical mechanics. At the other extreme are chaotic dynamical systems, some of which are illustrated. Asymptotic behavior of discrete parameter

Markov processes, which may be thought of as random dynamical systems, is studied in depth, including Harris recurrence. Standing apart are monotone Markov processes which may not even be irreducible, but still converge to equilibrium in special distances. The latter arise in the study of sustainable resources, queuing, etc. For functions of ergodic Markov processes, the Billingsley-Ibragimov FCLT is derived. A rather new exposition of the large deviation theory is presented covering both the classical i.i.d. case of Cramer and Sanov, as well as the Markovian case due to Donsker and Varadhan. There is a chapter on random fields which arise in physics that includes the FKG inequality and a central limit theorem under positive dependence due to Newman.

GTM 299 The contents of the final volume of the trilogy are summarized in the Preface.

Contents

Chapter 1
A Review of Martingales, Stopping Times, and the Markov Property

Primarily for ease of reference, this chapter contains an expository overview of some of the most essential aspects of martingale theory used in this text, especially as it pertains to continuous parameter Markov processes.

Among the dominant notions from the theory of stochastic processes used to develop the theory of stochastic differential equations are those of martingales, stopping times, and the Markov property. This chapter provides an expository account of this basic background material that will be used repeatedly in the ensuing applications to Markov processes and stochastic differential equations. A more exhaustive and detailed treatment is presented in BCPT.[1]

First, let us recall the meaning of the *Markov property* of a discrete parameter stochastic process $X = (X_0, X_1, \dots)$ defined on a probability space $(\Omega, \mathcal{F}, P)$, each X_n taking values in a measurable space $(S, \mathcal{S})$. Suppose that the initial state is deterministic, say $X_0 = x$ for some $x \in S$, and X has (time-homogeneous or stationary) transition probabilities $p(x, dy)$. Here $p(x, dy)$ is, for each $x \in S$, a probability measure on $(S, \mathcal{S})$ such that $x \to p(x, B)$ is measurable for every $B \in \mathcal{S}$. Then, the distribution of the process X is a probability measure P_x on the infinite product space $(S^\infty, \mathcal{S}^{\otimes\infty})$ of possible sample paths given by

$$P_{x_0}(A_0 \times A_1 \times \cdots \times A_n \times S^\infty) = P(\{\omega \in \Omega : X_i(\omega) \in A_i, 0 \le i \le n\} | X_0 = x_0), \tag{1.1}$$

[1] Throughout, BCPT refers to Bhattacharya and Waymire (2016), A Basic Course in Probability Theory.

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_1

for finite dimensional events $A_0 \times A_1 \times \cdots \times A_n \times S^\infty$, $A_i \in \mathcal{S}, 0 \le i \le n$. Let $\mathcal{F}_n = \sigma(X_0, X_1, \ldots, X_n)$ be the σ-field of subsets of Ω of this form, i.e., the σ-field generated by $X_0, X_1, \ldots, X_n$. In (1.1), the notation "$|X_0 = x_0$" is used to indicate conditional probability given $\sigma(X_0)$ evaluated at $X_0 = x_0$. Then, the Markov property may be expressed as follows: For each $m \ge 1$ (fixed), the conditional distribution of the *after-m process* $X_m^+ = (X_m, X_{m+1}, X_{m+2}, \ldots)$ given $\mathcal{F}_m$ is P_{X_m}, where P_{X_m} is the probability P_x with $x = X_m$. The Markov property may be expressed in terms of transition probabilities as: The conditional distribution of the state X_{m+1} given the past and present $\mathcal{F}_m$ is $p(X_m, dy)$. To see that this one-step Markov property is sufficient, observe that with it and iterated conditioning one has for any bounded measurable $f : S^{n+1} \to \mathbb{R}$,

$$\begin{aligned}
&\mathbb{E}\,(f(X_m, X_{m+1}, \ldots, X_{m+n})|\mathcal{F}_{m+n-1}) \\
&\quad= \int_S f(X_m, X_{m+1}, \ldots, X_{m+n-1}, x_{m+n})\, p(X_{m+n-1}, dx_{m+n}) \\
&\quad= g_{n-1}(X_m, \ldots, X_{m+n-1}),
\end{aligned} \tag{1.2}$$

for a measurable function $g_{n-1} : S^n \to \mathbb{R}$. Next conditioning on $\mathcal{F}_{m+n-2}$, one obtains a function $g_{n-2}(X_m, \ldots, X_{m+n-2})$. Continuing in this way one arrives at

$$\begin{aligned}
\mathbb{E}\,(f(X_m, X_{m+1}, \ldots, X_{m+n})|\mathcal{F}_m) &= \mathbb{E}\,(g_1(X_m, X_{m+1})|\mathcal{F}_m) \\
&= g_0(X_m).
\end{aligned} \tag{1.3}$$

On the other hand, a similar iteration (with $m = 0$) yields a computation of $\mathbb{E}_x f(X_0, \ldots, X_n)$. In this way one arrives at

$$\mathbb{E}(f(X_0, X_1, \ldots, X_m)|\sigma(X_0)) = g_0(X_0).$$

The consistent family of finite-dimensional distributions so obtained determines the distribution P_{x_0} for every initial state x_0. By integrating P_{x_0} with respect to a probability measure $\mu(dx_0)$ on $(S, \mathcal{S})$ one arrives at the distribution P_μ of the Markov process on $(S^\infty, \mathcal{S}^{\otimes\infty})$ having the transition probability $p(x, dy)$ and initial distribution $\mu(dx)$. In this notation $P_{x_0} \equiv P_{\delta_{x_0}}$.

An increasing sequence of subsigma-fields (of $\mathcal{F}$) $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}_m \subseteq \cdots$ is called a *filtration* on $(\Omega, \mathcal{F})$. If, in addition, one has that each X_n is $\mathcal{F}_n$-measurable, then we say that the process X is *adapted* to the filtration. This would be the case, for example, in the minimal case $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$. A *stopping time* with respect to the filtration is a nonnegative possibly infinite random variable τ defined on Ω, such that

$$[\tau \le n] \in \mathcal{F}_n, \quad \forall n \in \mathbf{Z}^+ = \{0, 1, 2, \ldots\} \cup \{\infty\}. \tag{1.4}$$

If τ is a stopping time, then the *pre-τ-sigmafield* may be defined by

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap [\tau \le n] \in \mathcal{F}_n \ \forall n \in \mathbf{Z}^+\}. \tag{1.5}$$

A typical example of a stopping time is the *first hitting time* of a set $B \in \mathcal{S}$ defined by

$$\tau_B = \inf\{n \ge 0 : X_n \in B\}. \tag{1.6}$$

The following proposition summarizes some simple and important properties.

Proposition 1.1 *(i) A nonnegative constant random variable τ is a stopping time. (ii) If τ_1 and τ_2 are stopping times, then so are $\tau_1 \wedge \tau_2 = \min\{\tau_1, \tau_2\}$, $\tau_1 \vee \tau_2 = \max\{\tau_1, \tau_2\}$, and $\tau_1 + \tau_2$. (iii) If in addition $\tau_1 \le \tau_2$, then $\mathcal{F}_{\tau_1} \subseteq \mathcal{F}_{\tau_2}$.*

The importance of stopping times for Markov processes is in the following extension of the Markov property to the *strong Markov property*: The conditional distribution of the *after-τ process* $X_\tau^+ = \{X_\tau, X_{\tau+1}, X_{\tau+2}, \dots\}$ given $\mathcal{F}_\tau$ is P_x evaluated at $x = X_\tau$, denoted $P_{X_\tau} = P_x|_{x=X_\tau}$, on the event $[\tau < \infty]$. This means precisely that

$$\mathbb{E}\left(\mathbf{1}_{A\cap[\tau<\infty]}\mathbf{1}_{[X_\tau^+\in B]}\right) = \mathbb{E}\left(\mathbf{1}_{A\cap[\tau<\infty]}P_{X_\tau}(B)\right), \tag{1.7}$$

for all $A \in \mathcal{F}_\tau$ and $B \in \mathcal{S}^{\otimes\infty}$.

Proposition 1.2 *If $X = (X_0, X_1, \dots)$ is a discrete parameter Markov process, then the strong Markov property also holds.*

Proof Assume that the initial state is $X_0 = x$. Let $A \cap [\tau = m] = [(X_0, X_1, \dots, X_m) \in C_m]$ for some measurable subset C_m of S^{m+1}. The right side of (1.7) is $\sum_{m=0}^{\infty} \mathbb{E}_x\{\mathbf{1}_{[(X_0,X_1,\dots,X_m)\in C_m]}P_{X_m}(B)\}$, where the subscript $\mathbb{E}_x$ indicates expected value with respect to P_x. Now the left side equals, by the Markov property after conditioning on $\mathcal{F}_m$ for the mth term of the sum,

$$\sum_{m=0}^{\infty} P_x((X_0, X_1, \dots, X_m) \in C_m), X_m^+ \in B) = \sum_{m=0}^{\infty} \mathbb{E}_x(\mathbf{1}_{[(X_0,X_1,\dots,X_m)\in C_m]}P_{X_m}(B)),$$

as asserted. ■

For homogeneous continuous parameter Markov processes $\{X_t : 0 \le t < \infty\}$, which is the subject matter of this book, one denotes the conditional distribution of X_{t+s}, given X_s, by $p(t; x, dy)$. But this family of transition probabilities cannot be chosen arbitrarily. In particular, these must satisfy the so-called *Chapman–Kolmogorov equation*:

$$p(t+s; x, B) = \int_S p(s; x, dy)p(t; y, B) \text{ for all } s > 0, t > 0, \quad B \in \mathcal{S}, \tag{1.8}$$

where $\mathcal{S}$ is the Borel σ-field of the Polish state space S. Note that the right side integrates $P_x(X_{t+s} \in B | X_s = y) = p(t; y, B)$ over the distribution $p(s; x, dy)$ of X_s, given $X_0 = x$. It is simple to see from this that if $p(t; x, dy)$ is known for $t \leq h$, for some $h > 0$, however small, then the entire family can be constructed. That is, one needs some infinitesimal condition for the transition probability, such as its "derivative" at $t = 0$, since the initial condition is given: $p(0; x, dy) = \delta_x(dy)$. This is explained in detail in the next chapter. Given the family $p(t; x, dy)$, $t > 0$, $x \in S$, we consider the canonical construction of $\{X_t : t \geq 0\}$ on a suitable, but general, path space Ω. Such a path space is the set of all *càdlàg functions* on $[0, \infty)$, namely, functions ω with values in the metric space S which are right continuous, i.e., $\omega(t+) = \omega(t)$, and have left limit $\omega(t-)$, say, as $s \to t$ from right or left, respectively. The process is given by the coordinate projection $X_t = \omega(t)$. We will study the topology on the space $D([0, \infty) : S) \equiv D(0, \infty) := \Omega$ of càdlàg functions induced by the Skorokhod metric on it. First, consider $D([0, T])$, the space of càdlàg functions on $[0, T]$ for each $T > 0$. Let Λ be the set of all continuous strictly increasing functions λ on $[0, T]$ onto $[0, T]$. Define the metric d on $D([0, T])$ as

$$d(f, g) = \inf\{\varepsilon > 0 : \sup\{|f(t) - g(\lambda(t))| : 0 \leq t \leq T\} \leq \varepsilon\}, \tag{1.9}$$

for some λ satisfying

$$\sup\{|\lambda(t) - t| : 0 \leq t \leq T\} \leq \varepsilon. \tag{1.10}$$

To check that d is a metric note that (i) $d(f, f) = 0$, taking $\lambda(t) = t$ for all t; conversely, if $f \neq g$, then there exists a t such that either $f(t) > g(t) + \delta'$ on $[t, t+\delta]$ or $f(t) < g(t)+\delta'$ on $[t, t+\delta]$, for some $\delta' > 0, \delta > 0$. Suppose, if possible, $d(f, g) = 0$. This contradicts the existence of a sequence λ_n converging uniformly to the identity function $\lambda(s) = s$ and $g(\lambda_n(s))$ converging to $f(s)$ uniformly on $[t, t + \delta]$. Next, (ii) $d(f, g) = d(g, f)$ follows by the equalities

$$\sup\{|f(t) - g(\lambda(t))| : 0 \leq t \leq T\} = \sup\{|f(\lambda^{-1}(t) - g(t)| : 0 \leq t \leq T\},$$

and

$$\sup\{|\lambda(t) - t| : 0 \leq t \leq T\} = \sup\{|\lambda^{-1}(t) - t| : 0 \leq t \leq T\}.$$

Finally, the triangle inequality follows from

$$|f(t) - h(\gamma \circ \eta(t))| \leq |f(t) - g(\eta(t))| + |g(\eta(t)) - h(\gamma \circ \eta(t))|,$$

and

$$|\gamma \circ \eta(t) - t| \leq |\gamma \circ \eta(t) - \eta(t)| + |\eta(t) - t| \text{ for all } \gamma, \eta \in \Lambda, t \geq 0.$$

Define a modulus w on $D([0, T])$ as follows: For $0 \leq t, s > 0, s + t < T$, $f \in D([0, T])$,

$$w(f : [t, t + s)) = \sup\{|f(u) - f(v)| : t \leq u < v < t + s\}.$$

The following lemma[2] plays an important role in deriving properties of the Skorokhod topology.

Lemma 1 *Let $f \in D([0, T])$. Then, for every $\varepsilon > 0$, there exist k and $0 = t_0 < t_1 < \cdots < t_k = T$ such that*

$$w(f : [t_i, t_{i+1})) < \varepsilon \text{ for all } i = 0, \ldots, k - 1. \tag{1.11}$$

Proof Consider the supremum $\hat{s}$ of all s, $0 \leq s \leq T$, for which the assertion holds on $[0, s]$. Since f is right continuous at 0, $\hat{s} > 0$. Since $f(\hat{s}^-)$ exists, $[0, \hat{s}]$ can be decomposed with the above property. Suppose, if possible, $\hat{s} < T$. But the right continuity of f at $\hat{s}$ implies that the property holds on $[0, \hat{s} + \delta]$ for some $\delta > 0$, leading to a contradiction. ■

It follows from the lemma that for $f \in D([0, T])$, given any $\varepsilon > 0$, there are at most finitely many points t in $[0, T]$ such that $|f(t) - f(t-)| \geq \varepsilon$. Letting $\varepsilon = 1/n, (n = 1, 2, \ldots)$, a countable dense subset of $D([0, T])$ is the collection of those f, which take rational values at T and have constant rational values in intervals $[t_i(n), t_{i+1}(n))(i = 0, 1, \ldots, k(n))$ of the kind in the lemma. This leads to the following result.

Proposition 1.3 *(a) The space $D([0, T])$ with the Skorokhod topology induced by the metric d is a separable metric space. (b) If f is continuous at t and $f_n \to f$ in the metric d, then $f_n(t)$ converges to $f(t)$, and if f is continuous on $[0, T]$, then the convergence is uniform on $[0, T]$. (c) The relative topology of the set $C = C[0, T]$ of continuous functions on $[0, T]$ into S is that of uniform convergence on $[0, T]$. (d) Let $f \in D([0, T])$ and $f^{(n)} = f$ at the set of points $t(r) = \{t_i(r), i = 0, \ldots, k(r)\}, r = 1, 2, \ldots, n$, as above, and joined by steps between neighboring points s, t for $s \leq u < t$. Then, $f^{(n)} \to f$ in the metric d.*

Proof The proofs of (a),(d) follow from the steps to the separability argument above. Part (b) follows from the definition of the metric d, and part (c) is a consequence (b). ■

We now follow Billingsley's argument[3] to prove the following.

Proposition 1.4 *The projections $\Pi_{t(1),\ldots,t(k)}$ on $D([0, T])$ into S^k, defined by*

$$\Pi_{t(1),\ldots,t(k)}(f) = (f(t(1)), f(t(2)), \ldots, f(t(k)))$$

---

[2] See Billingsley (1968), p. 110.

[3] See Billingsley (1968), p.121.

are Borel measurable (from $D([0, T[)$ *into* $(S^k, \mathcal{S}^{\otimes k})$ *for all time points* $0 \le t(1) \le \cdots \le t(k) \le T$*, for all* k*.*

Proof Π_0 is continuous (on $D([0, T])$ into S) due to right continuity at 0, and Π_T is continuous at T by definition of d. It is enough to prove Π_t is measurable for every t, $0 < t < T$. Fix such a t. Let $f_n \to f$ in the metric d. Then, $f_n \to f$ at all points of continuity of f, i.e., all points outside an at most countable set. Also, f_n are uniformly bounded. Therefore, as $n \to \infty$,

$$\frac{1}{\delta}\int_{[t,t+\delta]} f_n(s)ds \to \frac{1}{\delta}\int_{[t,t+\delta]} f(s)ds, \quad \text{for all } \delta > 0, (t + \delta < T).$$

This implies that $\frac{1}{\delta}\int_{[t,t+\delta]} f(s)ds$ is a continuous real-valued function in the Skorokhod topology and, by right-continuity of f at t, it converges to $f(t)$ as $\delta \downarrow 0$. This proves the measurability of Π_t. ■

The space $D([0, T[)$ is Polish under the Skorokhod topology. It is not complete under the metric d, but the topology under d is metrizable with a metric d_0 under which $D([0, T])$ is complete.[4] One defines the space $D([0, \infty))$ of càdlàg functions on $[0, \infty)$ into a Polish space S, with a metric

$$d(f, g) = \sum_{1\le n<\infty} 2^{-n}\frac{d_n(f, g)}{1 + d_n(f, g)},$$

where d_n is the metric d on $D([0, n])$ as defined above for $T = n$.

We now continue with the Markov process

$$X_t(\omega) = \Pi_t(\omega),\ 0 \le t < \infty,\ \omega \in \Omega = D([0, \infty)),$$

with values in a Polish space S. Let P_x denote its distribution when the initial state is x. For simplicity, we first consider $\Omega = \Omega_T = D([0, T])$, with corresponding expectations denoted $\mathbb{E}_x$.

Definition 1.1 The Markov process, and its transition probabilities $p(t; x, dy)$, are said to have the *Feller property*, also called the *weak Feller property*, if $x \to p(t; x, dy)$ is continuous in the weak topology of probability measures on $(S, \mathcal{S})$, i.e., for every real-valued bounded continuous function φ on S, $x \to T_t\varphi(x) = \int_S \varphi(y)p(t; x, dy)$ is continuous on S, for all $t > 0$.

Let us now show that the finite dimensional distributions also have the Feller property. For this, let $g(y_1, \ldots, y_k)$ be a real-valued bounded continuous function on S^k, and let $0 < t_1 < \cdots < t_k \le T$. Denote $\mathcal{F}_t = \sigma(X_s : 0 \le s \le t), t > 0$. With

[4] See Billingsley (1968), pp. 113–116.

$X_0 = x$,

$$\begin{aligned}
&\mathbb{E}_x \varphi(X_{t_1}, \dots, X_{t_k}) \\
&\quad = \mathbb{E}_x \mathbb{E}(\varphi(X_{t_1}, \dots, X_{t_k}) | \mathcal{F}_{t_{k-1}})) \\
&\quad = \mathbb{E}_x \int_S \varphi(X_{t_1}, \dots, X_{t_{k-1}}, y) p(t_k - t_{k-1}; X_{t_{k-1}}, dy_k) \\
&\quad = \mathbb{E}_x \phi_1(X_{t_1}, \dots, X_{t_{k-1}}),
\end{aligned}$$

where

$$\varphi_1(y_1, \dots, y_{k-1}) = \int_S \varphi(y_1, \dots y_{k-1}, y_k) p(t_k - t_{k-1}; y_{k-1}, dy_k)$$

at $X_{t_i} = y_i, i = 1, \dots, k-1$, which is a continuous function of y_{k-1}, for given $y_1, \dots, y_{k-2}$. But it is also continuous in the remaining variables $y_1, \dots, y_{k-2}$, by the continuity property of integration. Repeating this one arrives at

$$\mathbb{E}_x \varphi(X_{t_1}, \dots, X_{t_k}) = \mathbb{E}_x \phi_{k-1}(X_{t_1}) = \int_S \varphi_{k-1}(y_1) p(t_1; x, dy_1), \tag{1.12}$$

where φ_{k-1} is a continuous function on S:

$$\varphi_{k-1}(y_1) = \int_S \varphi_{k-2}(y_1, y_2) p(t_2 - t_1; y_1, dy_2).$$

This proves the assertion that finite-dimensional distributions under P_x are weakly continuous in x. We are now ready to prove the following result.

Proposition 1.5 *If the transition probabilities $p(t; x, dy), t > 0$, have the Feller property, then $x \to P_x$ is weakly continuous on the Polish space S into the Polish space of all probability measures on $D([0, T])$ and on $D([0, \infty))$.*

Proof Let ϕ be a bounded real-valued continuous function on $D([0, T])$. By Proposition 1.3(d), there is a countable set of points in $[0, T]$ such that every $f \in D([0, T])$ is a limit of functions $f^{(n)}$, as $n \to \infty$, whose values depend on the projection of $D([0, T])$ onto a finite set of points in $[0, T]$, i.e., $f^{(n)}$ are finite dimensional. Choose and fix $\varepsilon > 0$. Let $n(\varepsilon)$ be such that $\mathbb{E}_x \phi = \int \phi(f) dP_x$ differs from $\mathbb{E}_x \phi_n \equiv \int \phi(f^{(n)}) dP_x$ by no more than $\varepsilon/2$ for $n \geq n(\varepsilon)$. Let $x^{(k)} \to x$ as $k \to \infty$. Choose $k = k(\varepsilon)$ so that $\mathbb{E}_{x^{(k)}} \phi(n(\varepsilon))$ differs from $\mathbb{E}_x \phi(n(\varepsilon))$ by no more than $\varepsilon/2$ for $k \geq k(\varepsilon)$. Then, $\mathbb{E}_{x^{(k)}} \phi$ differs from $\mathbb{E}_x \phi$ by no more than ε for all $k \geq k(\varepsilon)$. This argument extends to P_x on $D([0, \infty))$ in view of the definition of the metric

$$d(f, g) = \sum_{1 \leq n < \infty} 2^{-n} d_n(f, g), \; f, g \in D[0, \infty).$$

■

A *continuous time filtration* is a family $\{\mathcal{F}_t : t \geq 0\}$ of subsigma-fields of $\mathcal{F}$, such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ if $s \leq t$. A *stopping time* with respect to this filtration is a nonnegative random variable $\tau : \Omega \to [0, \infty]$ such that

$$[\tau \leq t] \in \mathcal{F}_t \quad \text{for all } t \geq 0. \tag{1.13}$$

Theorem 1.6 (Strong Markov Property) *Suppose that $X = \{X_t : t \geq 0\}$ is a Markov process such that X_t takes values in a locally compact metric space S and has right-continuous sample paths. If the transition probabilities $p(t; x, dy)$ have the (weak) Feller property, then X has the strong Markov property. Namely, for any stopping time $\tau : \Omega \to [0, \infty]$, $A \in \mathcal{F}_\tau$, $B \in \mathcal{A}$, $x \in S$, one has*

$$\mathbb{E}\left(\mathbf{1}_{A\cap[\tau<\infty]}\mathbf{1}_{[(X_\tau^+ \in B]}\right) = \mathbb{E}\left(\mathbf{1}_{A\cap[\tau<\infty]}P_{X_\tau}(B)\right) \tag{1.14}$$

Proof The idea is to discretize time to get the Markov property and, hence, the strong Markov property there. Then, use the regularity conditions to lift it up to continuous time. Specifically, let D_n be the set of dyadic numbers $\{k2^{-n} : k = 0, 1, 2\ldots\}$, and $D_\infty = \cup_{n=1}^\infty D_n$, the set of all dyadic numbers. Discretely approximate τ by $\tau^{(n)} = (k+1)2^{-n}$ on $[k2^{-n} \leq \tau < (k+1)2^{-n}]$, for $k = 0, 1, 2, \ldots$. Then, $\tau^{(n)} \downarrow \tau$ as $n \uparrow \infty$. Letting $\mathcal{F}_{k,n} = \mathcal{F}_{k2^{-n}}, k = 0, 1, \ldots,$ note that $\tau^{(n)}$ is a stopping time with respect to both filtrations, $\mathcal{F}_t, t \geq 0$ and $\mathcal{F}_{k,n}, k = 0, 1, 2, \ldots$. Moreover, for each n, $\{X_{k,n} = X_{k2^{-n}} : k = 0, 1, 2, \ldots\}$ is a discrete parameter Markov process adapted to $\mathcal{F}_{k,n}, k = 0, 1, 2, \ldots$. In particular, therefore by Proposition 1.2, $\{X_{k,n} = X_{k2^{-n}} : k = 0, 1, 2, \ldots\}$ has the strong Markov property. Now let $A \in \mathcal{F}_\tau^{(m)} = \mathcal{F}_{\tau,m}$, regarded as the pre-$\tau^{(m)}$ σ-field for the Markov process indexed by D_m. Let f be a bounded continuous functon on S^r, and let $0 < d_1 < d_2 < \cdots < d_r$ be r ordered arbitrary dyadics from D_∞. For each n, consider the random variables $Z = f(X_{d_1}, \ldots, X_{d_r})$ and $Z_n = f(X_{\tau^{(n)}+d_1}, \ldots, X_{\tau^{(n)}+d_r})$. Using the fact that $\mathcal{F}_\tau^{(m)} \subseteq \mathcal{F}_\tau^{(n)}$ for all $n \geq m$ one gets, by Proposition 1.2,

$$\mathbb{E}_x(Z_n\mathbf{1}_{A\cap[\tau^{(n)}<\infty]}) = \mathbb{E}_x(\mathbf{1}_{A\cap[\tau^{(n)}<\infty]}\mathbb{E}_{X_{\tau^{(n)}}}Z). \tag{1.15}$$

Letting $n \to \infty$, noting that by the (weak) Feller property $P_{X_{\tau^{(n)}}}$ converges weakly to P_{X_τ}, one obtains

$$\mathbb{E}_x\left(\mathbf{1}_{A\cap[\tau<\infty]}f(X_{\tau+d_1}, \ldots, X_{\tau+d_r})\right) = \mathbb{E}_x\left(\mathbf{1}_{A\cap[\tau<\infty]}\mathbb{E}_x f(X_{d_1}, \ldots, X_{d_r})|_{x=X_\tau}\right). \tag{1.16}$$

Since this holds for (i) all $A \in \mathcal{F}_\tau^{(m)}$ for all m and (ii) for all r and, for each r, all bounded continuous functions f on S^r, with arbitrary dyadic numbers $0 \leq d_1 < d_2 < \cdots < d_r$, the assertion holds. ■

Example 1 (Brownian Motion) A basic example is that of Brownian motion. The precise mathematical definitions may be analogously formulated in defining general

multivariate normal distributions starting from a single one-dimensional standard normal random variable. Let us recall the following:

Definition 1.2 Let $(\Omega, \mathcal{F}, P)$ be a probability space.

a. *One-dimensional standard Brownian motion* starting at zero is the stochastic process $B = \{B_t : t \geq 0\}$ on $(\Omega, \mathcal{F}, P)$ with values in $S = \mathbb{R}$, $B_0 = 0$, having independent Gaussian increments $B_t - B_s$, $0 < s < t$ with zero means and variances $t-s$, and such that the sample paths $t \to B_t(\omega)$, $\omega \in \Omega$ are continuous, i.e., the function $t \to B_t(\omega), t \geq 0$, is an element of $\mathbf{K} := C([0, \infty) : \mathbb{R})$. The process defined by $B_t^x = x + B_t, t \geq 0$, is one-dimensional standard Brownian motion starting at $x \in \mathbb{R}$. The distributon P_x of B^x is referred to as one-dimensional *Wiener measure* on the space of continuous paths starting at x.

b. For an integer $k \geq 1$, the *k-dimensional standard Brownian motion* starting at $\mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{R}^k$ is the k-dimensional vector $\mathbf{B}^{\mathbf{x}} = (B^{x_1}, \ldots, B^{x_k})$ of independent one-dimensional standard Brownian motions B^{x_i} starting at x_i, $1 \leq i \leq k$. The distribution $P_{\mathbf{x}}$ of $\mathbf{B}^{\mathbf{x}}$ is referred to as k-dimensional *Wiener measure* starting at $\mathbf{x} \in S = \mathbb{R}^k$.

c. Given a $k \times k$ matrix $\boldsymbol{\sigma}$ and a vector $\boldsymbol{\mu}$, the stochastic process $\mathbf{X}_t^{\mathbf{x}} = \mathbf{x} + \boldsymbol{\mu} t + \boldsymbol{\sigma} \mathbf{B}$, $t \geq 0$, is referred to as *k-dimensional Brownian motion* starting at $\mathbf{x} \in \mathbb{R}^k$ with *drift* coefficient $\boldsymbol{\mu}$ and *diffusion coeficient* $\mathbf{D} = \boldsymbol{\sigma}\boldsymbol{\sigma}'$, where the superscript $'$ denotes matrix transpose.

In general, it is convenient to allow the matrix $\boldsymbol{\sigma}$ to be *singular*. In particular, in the one-dimensional case with $\sigma = 0$, $X_t^x = x + \mu t$ has the degenerate (singular) normal distribution with mean $x + \mu t$ and variance zero. Similarly if $\boldsymbol{\sigma}$ is of rank smaller than k, one obtains a singular normal distribution for the increments. The matrix $\mathbf{D} = \boldsymbol{\sigma}\boldsymbol{\sigma}'$ is a symmetric, nonnegative definite matrix, and it is customary to specify $\mathbf{D}$, with the convention that $\boldsymbol{\sigma}$ is the nonnegative definite matrix square root $\mathbf{D}^{\frac{1}{2}}$.

In view of their stationary, independent increments, the k-dimensional Brownian motions have the Markov property with homogeneous (possibly singular) Gaussian transition probabilities. In particular, the transition probabilities have the (weak) Feller property. This together with the sample path continuity implies that k-dimensional Brownian motion is a strong Markov process.

In the case $\boldsymbol{\sigma}$ is non-singular, the transition probabilities $p(t; \mathbf{x}, d\mathbf{y})$ are absolutely continuous with respect to Lebesgue measure and have the densities

$$p(t; \mathbf{x}, \mathbf{y}) = \frac{1}{(\sqrt{2\pi})^k \sqrt{\det \mathbf{D} t}} e^{-\frac{1}{2t}(\mathbf{y}-\mathbf{x}-\boldsymbol{\mu} t)\boldsymbol{D}^{-1}(\mathbf{y}-\mathbf{x}-\boldsymbol{\mu} t)'}, \ \mathbf{x}, \mathbf{y} \in S = \mathbb{R}^k, t > 0. \tag{1.17}$$

Example 2 (Poisson Process) The second important example that frequently occurs in the general theory is the Poisson process. The Poisson distribution $\frac{\nu^k}{k!} e^{-\nu}$, $k = 0, 1, \ldots$, is typically introduced as "law of rare event" approximating distribution

to the binomial distribution $\binom{n}{k}p^k(1-p)^{n-k}, k = 0, 1, \ldots, n$ in the limit as $p \downarrow 0, n \to \infty, np = \nu > 0$. In fact, using the simple *coupling inequality*

$$|P(S \in A) - P(T \in A)| \le P(S \ne T), \tag{1.18}$$

one can prove (see Exercise 1).

Lemma 2 (Le Cam's Poisson Approximation by Coupling[5]) *If $X_1, \ldots, X_n$ are independent $0-1$-valued Bernoulli random variables with $P(X_j = 1) = p_j \in (0, 1)$, then for $\nu = \sum_{j=1}^n p_j$,*

$$|P(\sum_{j=1}^n X_j = k) - \frac{\nu^k}{k!}e^{-\nu}| \le \sum_{j=1}^n p_j^2 \le \max_{1\le j\le n} p_j \sum_{j=1}^n p_j,\ k = 0, 1, \ldots n.$$

In particular, for $\nu = np$,

$$|\binom{n}{k}p^k(1-p)^{n-k} - \frac{\nu^k}{k!}e^{-\nu}| \le np^2. \tag{1.19}$$

While there are many ways in which to build on such approximations to introduce the Poisson processes, the following is a characteristic representation which we take as definition.

Definition 1.3 Let $T_0 = 0$ and $T_1, T_2, \ldots$ be i.i.d. exponentially distributed random variables with parameter $\lambda > 0$ defined on a probability space $(\Omega, \mathcal{F}, P)$. The Poisson process $\{X_t : t \ge 0\}$ with intensity parameter $\lambda > 0$ is the nonnegative integer valued stochastic process on $(\Omega, \mathcal{F}, P)$ defined by

$$X_t = \sup\{n \ge 0 : \sum_{j=0}^n T_j \le t\},\ X_0 = 0, \quad t \ge 0.$$

One may recall the argument that the (homogeneous) Poisson process X is a process with independent increments $X_t - X_s, 0 \le s \le t$, having Poisson distribution with parameter $\lambda(t-s)$ rests on first showing that given $[X_t = k]$, the conditional distribution of the arrival times $(A_1, \ldots, A_k)$ coincides with the distribution of the order statistic of k i.i.d. random variables with distribution function $F_t(s), 0 \le s \le t$, where, in the case of the Poisson process one finds that $F_t(s) = \frac{s}{t}, 0 \le s \le t$, is the uniform distribution on $[0, t]$. This is established by simply noting that the joint pdf of the i.i.d. inter arrival times is a product of exponentials and the arrival times are a unit Jacobian linear transformation of these. In particular, therefore, the conditional distribution of the first $k+1$ arrival times is obtained by normalizing

[5] See Bhattacharya and Waymire (2021), Proposition 5.1.

the latter joint pdf by $P(X_t = k) = P(A_k \le t < A_{k+1}) = \frac{\lambda^k t^k}{k!} e^{-\lambda t}$; this being a direct calculation from the Gamma distribution of the arrival times as sums of i.i.d. exponentials. Integrating the $k+1$-st arrival time of the normalized density over $[t, \infty)$ yields $k! \frac{1}{t^k}$. The following expands on this property in an interesting way.

Definition 1.4 (Order Statistic Property) Let $X = \{X_t : t \ge 0\}$ be a stochastic process whose sample paths are a.s. increasing unit step functions with jumps at $\beta_1, \beta_2, \dots$. Then, X is said to have the *order statistic property* (o.s.p.) if conditionally given $[X_t = n]$, $\beta_1, \dots, \beta_n$ are distributed as the order statistic of n i.i.d. random variables on $[0, t]$ with distribution function F_t.

Although one may prove the Markov property for the Poisson process directly from the independence of increments, the following theorem provides a broader perspective (also see Exercise 3 (b) for an application).

Theorem 1.7 (Order Statistic Property & Markov Property[6]) *Suppose that $X = \{X_t : t \ge 0\}, X_0 = 0$, is a stochastic process with (nonconstant) right-continuous increasing paths with unit jumps having the order statistic property. Then, X has the Markov property.*

Example 3 (Hitting Times as Stopping Times) The stopping time property (1.6) of hitting times for discrete parameter processes extends[7] to continuous time under some further regularity.

Proposition 1.8 *Suppose that the stochastic process $X = \{X_t : t \ge 0\}$, taking values in a metric space S with Borel σ-field $\mathcal{S} = \mathcal{B}$, has continuous sample paths. If B is a closed subset of S, then*

$$\tau_B = \inf\{t \ge 0 : X_t \in B\} \tag{1.20}$$

defines a stopping time with respect to the filtration $\mathcal{F}_t = \sigma(X_s : s \le t)$.

For the highlights of martingale theory that will be used in this text, let us first review the notion of a *discrete parameter martingale* with respect to a filtration $\mathcal{F}_n, n \ge 0$, as a real-valued stochastic process $X = \{X_0, X_1, X_2, \dots\}$, defined on a probability space $(\Omega, \mathcal{F}, P)$ and adapted to the filtration, such that $\mathbb{E}|X_n| < \infty$ for each n, with the property that

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n, \ n = 0, 1, 2, \dots. \tag{1.21}$$

A standard example is the simple symmetric random walk on $S = \mathbf{Z}$.

A *continuous parameter martingale* with respect to a filtration $\mathcal{F}_t, t \ge 0$, is a real-valued stochastic process $X = \{X_t : t \ge 0\}$, defined on a probability space

---

[6] This result is due to Crump (1975). Also see Bhattacharya and Waymire (2021), Proposition 5.3.

[7] See BCPT, p.59.

$(\Omega, \mathcal{F}, P)$ and adapted to the filtration, such that $\mathbb{E}|X_t| < \infty$ for each $t \geq 0$ and

$$\mathbb{E}(X_t|\mathcal{F}_s) = X_s,\ 0 \leq s \leq t. \tag{1.22}$$

A standard example is one-dimensional Brownian motion with drift parameter $\mu = 0$.

Proposition 1.9 [8] *(a) Let Y be integrable and $\mathcal{F}_n (n = 1, 2, ...)$ a filtration. Then, the martingale $Y_n = \mathbb{E}(Y|\mathcal{F}_n)$ is uniformly integrable. (b) Suppose Y_n is a $\mathcal{F}_n$-martingale $(n = 1, 2, ...)$ such that $Y_n \to Y$ in L^1. Then, $Y_n = \mathbb{E}(Y|\mathcal{F}_n), n \geq 1$.*

Remark 1.1 One can show that a $\mathcal{F}_n$-martingale $\{Z_n : n \geq 1\}$ has the representation $Z_n = \mathbb{E}(Z|\mathcal{F}_n)$ if and only if it is uniformly integrable, and then $Z_n \to Z$ a.s. and in L^1. Indeed, a uniformly integrable martingale converges a.s. and in L^1 by Doob's martingale convergence theorem.[9]

Theorem 1.10 (Discrete Parameter Optional Stopping[10]) *Suppose that $\{X_n : n = 0, 1, \ldots\}$ is a martingale with respect to a filtration $\mathcal{F}_n, n \geq 0$. (a) If $\tau_1 \leq \tau_2$ are bounded stopping times with respect to this filtration, then $\mathbb{E}(X_{\tau_2}|\mathcal{F}_{\tau_1}) = X_{\tau_1}$. (b) If $\tau_1 \leq \tau_2 < \infty$ a.s. are stopping times such that $\{X_{\tau_2 \wedge n} : n = 0, 1, 2, \ldots\}$ is a uniformly integrable collection of random variables, then $\mathbb{E}(X_{\tau_2}|\mathcal{F}_{\tau_1}) = X_{\tau_1}$. In particular, $\mathbb{E}X_{\tau_2} = \mathbb{E}X_{\tau_1} = \mathbb{E}X_0$.*

Remark 1.2 The conclusion of optional stopping may also be phrased as the fact that the two-term sequence X_{τ_1}, X_{τ_2} is a martingale with respect to the filtration pair $\mathcal{F}_{\tau_1}, \mathcal{F}_{\tau_2}$.

Corollary 1.11 (Optional Stopping[11]) *Suppose that $\{X_n : n = 0, 1, \ldots\}$ is a martingale and τ is a stopping time, each with respect to the filtration $\mathcal{F}_n, n \geq 0$. Then $\{X_{\tau \wedge n} : n = 0, 1, 2, \ldots\}$ is a martingale with respect to the filtration $\mathcal{F}_{\tau \wedge n}, n \geq 0$.*

Remark 1.3 It follows from Jensen's inequality for conditional expectation that if $\{X_n : n \geq 0\}$ is a martingale, then

$$\mathbb{E}(|X_{n+1}||\mathcal{F}_n) \geq |X_n|,\ n = 0, 1, 2, \ldots. \tag{1.23}$$

Accordingly, the stochastic process $|X_n|, n = 0, 1, 2, \ldots$ is said to be a *submartingale*. That is, the equality "=" defining martingales in (1.21) becomes an inequality "≥" in the definition of a submartingale. The optional stopping theorem extends to submartingales with the conclusion modified to $\mathbb{E}(X_{\tau_2}|\mathcal{F}_{\tau_1}) \geq X_{\tau_1}$.

---

[8] See BCPT, Proposition 3.7.

[9] See BCPT Theorems 3.11, 3.12.

[10] See BCPT, Proposition 3.8.

[11] See BCPT, Theorem 3.8.

Similarly, the corollary may be restated for submartingales with the conclusion that $\{X_{\tau\wedge n} : n \geq 0\}$ is a submartingale too.

The continuous parameter version again requires some sample path regularity.

Theorem 1.12 (Continuous Parameter Optional Stopping) *Suppose $X = \{X_t : t \geq 0\}$ is a right-continuous martingale with respect to a filtration $\mathcal{F}, t \geq 0$. If $\tau_1 \leq \tau_2$ are a.s. finite stopping times with respect to this filtration and if $\{X_{\tau_2\wedge t} : t \geq 0\}$ is a uniformly integrable collection of random variables, then $\mathbb{E}(X_{\tau_2}|\mathcal{F}_{\tau_1}) = X_{\tau_1}$, and $\mathbb{E}X_{\tau_2} = \mathbb{E}X_{\tau_1} = \mathbb{E}X_0$.*

Corollary 1.13 *Suppose that $\{X_t : t \geq 0\}$ is a right-continuous martingale, and τ a (possibly infinite) stopping time, each with respect to a filtration $\mathcal{F}_t, t \geq 0$. Then $\{X_{\tau\wedge t} : t \geq 0\}$ is a martingale with respect to $\mathcal{F}_{\tau\wedge t}, t \geq 0$. If $\{X_t : t \geq 0\}$ is a right-continuous submartingale, then $\{X_{\tau\wedge t} : t \geq 0\}$ is also a submartingale with respect to $\mathcal{F}_{\tau\wedge t}, t \geq 0$.*

We conclude this chapter with two useful lemmas on integration by parts. But we also include two of the most important inequalities in probability theory as well.

Theorem 1.14 (Doob's Maximal Inequality[12]) *Assume $\{X_1, X_2, \dots, X_n\}$ is an $\{\mathcal{F}_k : 1 \leq k \leq n\}$-martingale, or a nonnegative submartingale, and $\mathbb{E}|X_n|^p < \infty$ for some $p \geq 1$. Then, for all $\lambda > 0$, $M_n := \max\{|X_1|, \dots, |X_n|\}$ satisfies*

$$P(M_n \geq \lambda) \leq \frac{1}{\lambda^p}\int_{[M_n>\lambda]} |X_n|^p dP \leq \frac{1}{\lambda^p}\mathbb{E}|X_n|^p. \tag{1.24}$$

The following lemmas illustrate the often used exchange in the order of integration and will be used to prove Doob's second maximal inequality.

Lemma 3 (Integration by Parts[13]) *Let μ_1, μ_2 be signed measures on $\mathbb{R}$, which are finite on finite intervals. Let*

$$F_i(y) = \mu_i(0, y], \quad i = 1, 2, \quad -\infty < y < \infty.$$

Then, for any $-\infty < a < b < \infty$, one has

$$\int_{(a,b]} F_1(y)\mu_2(dy) = F_1(b)F_2(b) - F_1(a)F_2(a) - \int_{(a,b]} F_2(y^-)\mu_1(dy).$$

Remark 1.4 The "distribution functions" $F_i, i = 1, 2$ can be defined as $F_i(y) = \mu_i((c, y]), i = 1, 2, y \in \mathbb{R}$, for any real number c in place of zero, and the lemma still holds. This formula has special utility when applied to a nondecreasing

[12] See BCPT, Theorem 3.2.

[13] See BCPT, p.9.

function, or more generally a function of bounded variation, as an integrand, the latter being a difference of nondecreasing functions.

The following is a useful version[14] for expected values. The proof[15] of Theorem 1.15 makes a nice application.

Lemma 4 *Let μ_1 be an arbitrary measure on $(0, \infty]$, which is finite on finite intervals, and such that $\mu_1(\{0\}) = 0$. Suppose that μ_2 is a probability measure on $[0, \infty)$, and let Y be a random variable with distribution μ_2. Then, with $F_i(y) = \mu_i((0, y]), i = 1, 2, y \geq 0$, one has*

$$\mathbb{E}F_1(Y) = \int_{[0,\infty)} P(Y \geq y)\mu_1(dy)$$

Theorem 1.15 (Doob's Maximal Inequality for Moments) *Let $\{X_1, X_2, \ldots, X_n\}$ be an $\{\mathcal{F}_k : 1 \leq k \leq n\}$-martingale, or a nonnegative submartingale, and let $M_n = \max\{|X_1|, \ldots, |X_n|\}$. Then,*

a. $\mathbb{E}M_n \leq \frac{e}{e-1}\left(1 + \mathbb{E}|X_n| \log^+ |X_n|\right)$.
b. *If $\mathbb{E}|X_n|^p < \infty$ for some $p > 1$, then $\mathbb{E}M_n^p \leq q^p \mathbb{E}|X_n|^p$, where q is the conjugate exponent defined by $\frac{1}{q} + \frac{1}{p} = 1$, i.e., $q = \frac{p}{p-1}$.*

Corollary 1.16 *Let $\{X_t : t \in [0, T]\}$ be a right-continuous nonnegative $\{\mathcal{F}_t\}$-submartingale with $\mathbb{E}|X_T|^p < \infty$ for some $p \geq 1$. Then, $M_T := \sup\{X_s : 0 \leq s \leq T\}$ is $\mathcal{F}_T$-measurable and, for all $\lambda > 0$,*

$$P(M_T > \lambda) \leq \frac{1}{\lambda^p} \int_{[M_T > \lambda]} X_T^p dP \leq \frac{1}{\lambda^p} \mathbb{E}X_T^p, \ (p \geq 1), \tag{1.25}$$

and for all $p > 1$,

$$\mathbb{E}M_T^p \leq q^p \mathbb{E}X_T^p, \quad (q = \frac{p}{p-1}).$$

Proof Consider the nonnegative submartingale $\{X_{j2^{-n}T} : j = 0, 1, \ldots, 2^n\}$, for each $n = 1, 2, \ldots$, and let $M_n := \max\{X_{j2^{-n}T} : 0 \leq j \leq 2^n\}$. For $\lambda > 0$, $[M_n > \lambda] \uparrow [M_T > \lambda]$ as $n \uparrow \infty$. In particular, M_T is $\mathcal{F}_T$-measurable. By Theorem 1.14,

$$P(M_n > \lambda) \leq \frac{1}{\lambda^p} \int_{[M_n > \lambda]} X_T^p dP \leq \frac{1}{\lambda^p} \mathbb{E}X_T^p.$$

Letting $n \uparrow \infty$, (1.25) is obtained. The second inequality is obtained similarly. ■

[14] See BCPT, Proposition 1.4.

[15] See BCPT, Theorem 3.4.

For the next Remark, recall that $t \to X_t$ is *stochastically continuous* if $|X_t - X_s| \to 0$ in probability as $s \to t$, $(t \geq 0)$.

Remark 1.5 (Regularization of Continuous Parameter Submartingales) Let $\{X_t : t \geq 0\}$ be a submartingale such that $t \to X_t$ is stochastically continuous. Then there is a version $\{\widetilde{X}_t : t \geq 0\}$ of $\{X_t : t \geq 0\}$ having càdlàg sample paths almost surely (e.g., see Theorem 3.1, Chapter 3.)

Exercises

1. (i) Prove (1.18). (ii) Provide a proof of Le Cam's coupling approximation (1.19). [*Hint*: For i.i.d. uniformly distributed U_j, $j \geq 1$, on (0, 1), define

$$X_j = \begin{cases} 0 & 0 \leq U_j < 1 - p_j \\ 1 & 1 - p_j \leq U_j < 1, \end{cases}$$

and define

$$Y_j = \begin{cases} 0 & 0 \leq U_j < e^{-p_j} \\ 1 & e^{-p_j} \leq U_j < e^{-p_j} + p_j e^{-p_j} \\ 2 & e^{-p_j} + p_j e^{-p_j} \leq U_j < e^{-p_j} + p_j e^{-p_j} + \frac{p_j^2}{2!} e^{-p_j} \\ \vdots & \text{etc.} \end{cases}$$

2. Give a proof of Proposition 1.1.
3. (a) Prove Theorem 1.7. (b) *(Yule Process)* Let $\mathbb{T} = \cup_{n=0}^{\infty}\{1,2\}^n$, and view each $v = (v_1, \ldots, v_n) \in \mathbb{T}, n \geq 1$, as an nth generation vertex of height $|v| = n$, in a binary tree rooted at the zeroth generation vertex $\theta \in \{1,2\}^0$, $|\theta| = 0$. Next, let $\{T_v, v \in \mathbb{T}\}$ be a *random field* of i.i.d. exponential random variables with intensity parameter λ, indexed by $\mathbb{T}$, and consider the set $V_1(t) = \{v \in \mathbb{T} : \sum_{j=0}^{|v|-1} T_{v|j} \leq t < \sum_{j=0}^{|v|} T_{v|j}\}$, $\quad t \geq 0$, where $\sum_{j=0}^{-1} = 0$, and $v|j = (v_1, \ldots, v_j), v \in \mathbb{T}, j \leq |v|$. As the evolutionary set $V_1(t)$ evolves, when an exponentially distributed clock rings for a vertex v, then it is removed and two offspring vertices $v1$, $v2$ are incorporated. The stochastic process $Y_t = |V_1(t)|, t \geq 0$, counting the number of (offspring) vertices at time t, is referred to as a *Yule process* (compare Definition 1.3), starting at $Y_0 = |V_1(0)| = 1$, i.e., $V_1(0) = \{\theta\}$. Note that the sample paths of Y are nondecreasing step functions with unit jumps. Let $0 = \beta_0 < \beta_1 < \cdots$ denote the successive jump times of Y, i.e., $\beta_k = \inf\{t \geq \beta_{k-1} : Y_t = k+1\}$.

   (i) Show that for arbitrary $k \geq 1$, conditionally given $\beta_0 = 0$ and given $\sigma\{\beta_1, \ldots, \beta_{k-1}\}$, the increment $\beta_k - \beta_{k-1}$ is exponentially distributed with

parameter $k\lambda$. In particular, $\beta_k - \beta_{k-1}, k = 1, 2, \ldots$ are independent random variables. [*Hint*: $P(\beta_1 > t) = e^{-\lambda t}, t \geq 0$. Next, $P(\beta_2 - \beta_1 > t) = P(T^{(1)} \wedge T^{(2)} > t) = e^{-\lambda t}e^{-\lambda t} = e^{-2\lambda t}$, and more generally for $k \geq 2$, given $\beta_1, \ldots, \beta_{k-1}$, $\beta_k - \beta_{k-1}$ is the minimum of k i.i.d. exponentially distributed random variables with intensity λ.]

(ii) Show that for fixed $t \geq 0$, Y_t has a (marginal) geometric distribution with parameter $p_t = e^{-\lambda t}$. In particular, Y does not have stationary nor independent increments. [*Hint*: Compute the distribution of Y_t by a similar method as used for the Poisson distribution, i.e., write $P(Y_t = k) = P(\beta_k \leq t < \beta_{k+1})$ as a difference of probabilities and apply integration by parts.]

(iii) Show that Y has the o.s.p. with $F_t(s) = \frac{e^{\lambda s}-1}{e^{\lambda t}-1}, \quad 0 \leq s \leq t$. [*Hint*: Given $[Y_t = n + 1]$ there are n jump times $\beta_1 < \beta_2 < \cdots < \beta_n$. The joint pdf of $(\beta_1, \beta_2 - \beta_1, \ldots, \beta_{n+1} - \beta_n)$ is the product $(n + 1)!\lambda^{n+1}e^{-\lambda\sum_{j=1}^{n+1} jt_j}, t_j \geq 0$. Thus, by a change of variable, $(\beta_1, \ldots, \beta_{n+1})$ has pdf $(n + 1)!\lambda^{n+1}e^{-\lambda\sum_{j=1}^{n+1} j(s_j - s_{j-1})} = (n + 1)!\lambda^{n+1}e^{-\lambda\{(n+1)s_{n+1} - \sum_{j=1}^{n} s_j\}}, 0 = s_0 < s_1 < \cdots < s_{n+1}$. Normalize by $P(Y_t = n + 1)$ and integrate s_{n+1} over $[t, \infty)$ to obtain the desired conditional pdf $n!\prod_{j=1}^{n} \frac{\lambda e^{\lambda s_j}}{e^{\lambda t}-1}, 0 < s_1 < s_2 < \cdots < s_n \leq t$.]

(iv) Show that Y is a Markov process. [*Hint*: Apply Theorem 1.7 to $Y - 1$.]

(v) Show that the limit $W = \lim_{t\to\infty} e^{-\lambda t}Y_t$ exists a.s. and is exponentially distributed with mean one. [*Hint*: Check that $e^{-\lambda t}Y_t, t \geq 0$, is a uniformly integrable, nonnegative martingale. In particular, $\sup_t \mathbb{E}e^{-2\lambda t}Y_t^2 = \sup_t e^{-2\lambda t}\frac{2-e^{-\lambda t}}{e^{-2\lambda t}} < \infty$. Show that $W = e^{-\lambda\beta_1}(W^{(1)} + W^{(2)})$, in distribution, where $W^{(1)}$, $W^{(2)}$ are conditionally i.i.d. given β_1, and check that the unique positive, nonconstant, mean one solution to the corresponding equation for $\varphi(r) = \mathbb{E}e^{rW}$, namely, after conditioning on $T_\theta = \beta_1$, $\varphi(r) = \mathbb{E}\varphi^2(re^{-\lambda\beta_1}), \quad \varphi(0) = 1$, is $\varphi(r) = \frac{1}{1-r}$, $r < 1$. For this, make a change of variable, $\varphi(r) = \frac{1}{r}\int_0^r \varphi^2(u)du, \quad \varphi(0) = 1$. Differentiating it follows that $\varphi(r) = \frac{1}{1+cr}$, for a constant c. Finally, observe that $1 = \mathbb{E}W = \varphi'(0) = -c$.]

Chapter 2
Semigroup Theory and Markov Processes

Semigroups associated with Markov processes are introduced as examples together with some general theory and basic properties, including the Hille–Yosida theorem.

To construct a discrete parameter homogeneous Markov process with state space $(S, \mathcal{S})$, one only needs to specify the initial state, or initial distribution, and the one-step transition probability $p(x, dy)$ (see BCPT,[1] pp. 187–189, or Bhattacharya and Waymire 2021, pp. 9–11). There is, in general, no such simple mechanism in the continuous parameter cases to construct the transition probability $p(t; x, dy)$. This is because the latter transition probabilities must satisfy the following requirements (Exercise 5):

Definition 2.1 Let $(S, \mathcal{S})$ be a measurable space. A family of homogeneous continuous parameter transition probabilities consist of functions $p(t; \cdot, \cdot), t > 0$, on $S \times \mathcal{S}$ with the following properties:

(i) For each $t > 0$, each $x \in S$, $B \to p(t; x, B)$ is a probability measure on $\mathcal{S}$,
(ii) For each $t > 0$, each $B \in \mathcal{S}$, $x \to p(t; x, B)$ is a measurable function on S, into $\mathbb{R}$ on the σ-field $\mathcal{S}$ on S and the Borel σ-field $\mathcal{B}(\mathbb{R})$ on $\mathbb{R}$.
(iii) *(Chapman–Kolmogorov Equation)* For all $t > 0,\ s > 0,\ x \in S,\ B \in \mathcal{S}$,

[1] Throughout, BCPT refers to Bhattacharya and Waymire (2016), A Basic Course in Probability Theory.

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_2

$$p(t+s; x, B) = \int_S p(t; z, B) p(s; x, dz).$$

In view of the Chapman–Kolmogorov equation (iii) in Definition 2.1, one needs to determine the transition probability *only* on an arbitrarily small interval $(0, t_0)$. Indeed, in this chapter, we show that the derivative of the transition probability operator at $t = 0$ determines the transition probability for all times!

We begin with the primary example for this chapter, Markov semigroups, and then consider a number of specific examples to further elucidate the nature and properties of naturally ocurring semigroups.

Example 1 (Markov Semigroups) Let $(S, \mathcal{S})$ be a metric space with Borel σ-field $\mathcal{S}$. Let $p(t; x, dy)$ be a transition probability on S. Let $\mathbb{B}(S)$ denote the set of all real-valued bounded measurable functions on S. $\mathbb{B}(S)$ is a Banach space with respect to the *"sup"-norm*

$$||f|| = \sup_{x \in S} |f(x)|, \quad f \in \mathbb{B}(S). \tag{2.1}$$

Define the *transition operator* $T_t : \mathbb{B}(S) \to \mathbb{B}(S)$, by

$$(T_t f)(x) = \int_S f(y) p(t; x, dy), \quad t > 0. \tag{2.2}$$

By property (iii) of the transition probability $p(t; x, dy)$, $\{T_t : t > 0\}$ is a *one-parameter semigroup of (commuting) linear operators*:

$$T_{t+s} f = T_t(T_s f), \; f \in \mathbb{B}(S), \; \text{i.e.,} \; T_t T_s = T_s T_t. \tag{2.3}$$

Also, each T_t is a *contraction*:

$$||T_t f|| \le ||f||, \quad f \in \mathbb{B}(S), \tag{2.4}$$

as is easily seen from (2.2), $T_t f(x)$ being the average of f with respect to the probability measure $p(t; x, dy)$. Note that taking $f = \mathbf{1}_B, B \in \mathcal{S}$, implies the *Chapman–Kolmogorov equations* in the form

$$\begin{aligned} p(t+s; x, B) &= \int_S p(t; x, dy) p(s; y, B) \\ &= \int_S p(s; x, dy) p(t; y, B), s, t \ge 0, B \in \mathcal{S}. \end{aligned} \tag{2.5}$$

Example 2 (Initial Value Problem for a System of Linear Ordinary Differential Equations) Consider the system of linear equations

$$\frac{d}{dt}u(t) \equiv \dot{u}(t) = Au(t), \quad (u(t) = (u^{(1)}(t), u^{(2)}(t), \ldots, u^{(m)}(t))) \quad t \geq 0, \tag{2.6}$$

with *initial condition*

$$u(0) = f = (f^{(1)}, f^{(2)}, \ldots, f^{(m)}) \in \mathbb{R}^m, \tag{2.7}$$

and

$$A = ((q_{ij})), \text{ a constant (i.e., time-independent) } m \times m \text{ matrix.} \tag{2.8}$$

The *unique continuous solution* (on $[0, \infty)$]) to (2.6), (2.7) is

$$u(t) = e^{tA} f := \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n f, \tag{2.9}$$

as is easily checked by term-by-term differentiation. Writing

$$T_t f = e^{tA} f, \tag{2.10}$$

one has a one-parameter family $\{T_t : t > 0\}$ of linear operators on the finite-dimensional Banach space $\mathbb{R}^m$ satisfying the semigroup property

$$T_{t+s} = T_t T_s,$$

since $e^{(t+s)A} = e^{tA}e^{sA}$, as may be easily checked by expanding the three exponentials and using the binomial theorem. Equations such as (2.6) often arise as equations of motion for a dynamical system.

More generally, if $\mathscr{X}$ is a Banach space and A is a bounded linear operator on $\mathcal{D}_A = \mathscr{X}$ (into $\mathscr{X}$), then

$$T_t f \equiv e^{tA} f \quad (t > 0, \ f \in \mathscr{X}) \tag{2.11}$$

defines a one-parameter semigroup of bounded linear operators $\{T_t = e^{tA} : t > 0\}$, with $e^{tA} = \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n$. In particular, $u(t) \equiv T_t f$ is a solution of the initial value problem

$$\frac{d}{dt}u(t) = Au(t), \ u(0) = f, \quad t > 0. \tag{2.12}$$

If f is such that $T_t f \to f$ as $t \downarrow 0$ (in the Banach space norm), then $u(t) = T_t f$ is the *unique* solution to (2.12), which is continuous at $t = 0$ and satisfies $u(0) = f$.

The next example illustrates $u(t) = T_t f$ as a solution to an initial value problem in the context of a Markov chain on a countable state space S.

Example 3 (Markov Chains) First, let S be a finite set, say $S = \{1, 2, \dots, n\}$, and let $\mathcal{S}$ be the class of all subsets of S. Suppose $p(t; i, j)$ is a transition probability on S. Then,

$$(T_t f)(i) = \sum_{j \in S} f(j) p(t; i, j) \quad (i \in S). \tag{2.13}$$

Taking $f = f_j := \mathbf{1}_{\{j\}}$, the semigroup property implies the so-called *Chapman–Kolmogorov* equations for $p(t; i, j)$. Namely,

$$p(t+s; i, j) = \sum_k p(s; i, k) p(t; k, j) = \sum_k p(t; i, k) p(s; k, j), \quad i, j \in S, t \geq 0, \tag{2.14}$$

or as matrix products for the *transition probability matrix*:

$$p(t) = ((p(t; i, j)))_{ij \in S}, \;\; p(t+s) = p(s)p(t) = p(t)p(s). \tag{2.15}$$

Assume that for all i, j, the function $t \to p(t; i, j)$ is continuous (from the right) at $t = 0$. This is equivalent to the property that the semigroup T_t is strongly continuous on $C(S : \mathbb{R})$ (see Definition 2.6). In particular, for each pair (i, j), one has the finite limit

$$q_{ij} \equiv \lim_{t \downarrow 0} \frac{p(t; i, j) - p(0; i, j)}{t} = \lim_{t \downarrow 0} \frac{p(t; i, j) - \delta_{ij}}{t}, \tag{2.16}$$

so that by term-by-term differentiation at $t = 0$ of (2.13), one has

$$Af(i) = \sum_{j=1}^{n} q_{ij} f(j). \tag{2.17}$$

Since

$$\sum_{j=1}^{n} p(t; i, j) = 1 \quad (i \in S), \tag{2.18}$$

one also has

$$(a) \quad \sum_{j=1}^{n} q_{ij} = 0. \tag{2.19}$$

Also, since $p(t; i, i) \leq \delta_{ii} = 1$, $p(t; i, j) \geq \delta_{ij} = 0$ if $i \neq j$, (2.16) yields,

$$(b) \quad q_{ii} \leq 0, \quad q_{ij} \geq 0 \text{ if } i \neq j. \tag{2.20}$$

Thus, the matrix (operator) $A = ((q_{ij}))_{i,j \in S}$ has the following properties:

(i) *Nonpositive diagonals:* $q_{ii} \leq 0, i \in S$.
(ii) *Nonnegative off-diagonals:* $q_{ij} \geq 0, i \neq j, i, j \in S$.
(iii) *Zero row sums:* $\sum_{j \in S} q_{ij} = 0$.

By uniqueness of the solution

$$u(t) = e^{tA} f, \quad t \geq 0,$$

to the initial value problem

$$\dot{u}(t) = Au(t), \ t > 0, \quad u(0) = f, \tag{2.21}$$

one therefore has

$$T_t f = e^{tA} f, \quad (e^{tA} = \sum_{m=0}^{\infty} \frac{t^m}{m!} A^r). \tag{2.22}$$

In particular, letting f_j be the *indicator* of $\{j\}$,

$$p(t; i, j) = T_t f_j(i) = \left(e^{tA}\right)_{(i,j)}. \tag{2.23}$$

Conversely:

Proposition 2.1 *Let $S = \{1, 2, \ldots, n\}$. Given any $n \times n$* matrix *$A = ((q_{ij}))_{i,j \in S}$ satisfying (i)–(iii) above, (2.23) defines a semigroup of transition probabilities $p(t; i, j)$ such that $p'(0^+; i, j) = q_{ij}, i, j \in S$.*

Proof Note that (2.20) implies

$$p(t; i, j) \geq 0 \text{ for all } i, j \in S,$$

and

$$p(t; i, i) \leq 1, \text{ for all } i \in S, \text{ for all } t > 0.$$

On the other hand, letting $f \equiv 1$, one has

$$\sum_{j \in S} p(t; i, j) = (T_t f)(i). \tag{2.24}$$

But $T_t f \equiv 1$ solves (2.21) for $u(0) \equiv 1$ (in view of (2.19)). Hence, by *uniqueness* of the solution to (2.21),

$$\sum_{j \in S} p(t; i, j) = (T_t 1)(i) = 1. \tag{2.25}$$

By specializing (2.21) to the functions $f_j = \mathbf{1}_{\{j\}}$, (or using (2.23) directly), one gets

$$p(t; i, j) = \begin{cases} tq_{ij} + o(t) & \text{as } t \downarrow 0, \ \ i \neq j, \\ 1 - tq_{ii} + o(t) & \text{as } t \downarrow 0, \ \ i = j. \end{cases} \tag{2.26}$$

This completes the proof. ■

Remark 2.1 In this context, the quantity q_{ij} is referred to as the *rate of transition* from state i to state j $(i \neq j)$; of course q_{ii} is determined by q_{ij} $(j \neq i)$ since

$$q_{ii} = - \sum_{\{j : j \neq i\}} q_{ij}.$$

The parameters $q_{ij}, i \neq j$, and $-q_{ii}, i, j \in S$ will also be referred to as the *infinitesimal parameters (or infinitesimal rates)*. A more illuminating description of q_{ij}'s will be given later.

Next, assume that S is *denumerable* (i.e., infinite but countable). Let $A = ((q_{ij}))$ be a matrix satisfying properties (i)–(iii). This implicitly requires that

$$\sum_{\{j : j \neq i\}} q_{ij} < \infty \quad \text{for all } i \in S. \tag{2.27}$$

Then, it is easily checked that

Theorem 2.2 (Bounded Rates Theorem) *Under the conditions (i)–(iii) on* $A = ((q_{ij}))_{i,j \in S}$, *one has that*

$$Af(i) = \sum_{j \in S} q_{ij} f(j), \ \ i \in S, \ \ f \in \mathbb{B}(S),$$

defines a bounded linear operator on $\mathbb{B}(S)$ if and only if

$$\sup\{|q_{ii}| : i \in S\} < \infty. \tag{2.28}$$

If (2.28) holds, then

$$T_t f \equiv e^{At} f := \sum_{n=0}^{\infty} \frac{t^n A^n f}{n!}, \quad f \in \mathbb{B}(S),$$

defines a Markov semigroup whose transition probabilities are given by

$$p(t; i, j) = (e^{tA})_{(i,j)}. \tag{2.29}$$

The proof of this theorem is entirely analogous to that of the finite (state space) case and left as an exercise.

Remark 2.2 In case the matrix (operator) A is *unbounded* (i.e., (2.28) fails), then it may be shown that there exists at least one *sub-Markovian* transition probability (i.e., $\sum_j p(t; i, j) \leq 1$) satisfying (2.26). However, uniquenss is a more delicate affair (to be treated in Chapter 6).

Our program in this chapter is to show that, under appropriate conditions, one-parameter semigroups on a Banach space are of the form (2.10), suitably defined and interpreted to allow unbounded operators A as well. Before pursuing this program, let us consider two more examples.

Example 4 (Initial Value Problem for Parabolic PDE's) The equation of heat conduction (or diffusion of a solute) in a homogeneous medium (without boundary) is

$$\frac{\partial u(t,x)}{\partial t} = D\Delta_x u \quad (t > 0,\ x \in \mathbb{R}^k), \tag{2.30}$$

with the *initial condition*

$$u(0,x) = f(x). \tag{2.31}$$

Here $u(t,x)$ is the *temperature* in the medium *at time* t *at a point* x (In the physical problem, $k = 3$), $f(x)$ is the *initial temperature at this point*, and D is a positive constant. Here Δ_x is the *Laplacian*

$$\Delta_x = \sum_{j=1}^{k} \frac{\partial^2}{\partial x_j^2}. \tag{2.32}$$

By taking Fourier transforms of both sides (as functions of x, for fixed t) of (2.30), one arrives at

$$\frac{\partial \hat{u}(t,\xi)}{\partial t} = -D|\xi|^2 \hat{u}(t,\xi) \quad (t > 0,\ \xi \in \mathbb{R}^k) \tag{2.33}$$

whose general solution is

$$\hat{u}(t,\xi) = c(\xi)e^{-tD|\xi|^2}. \tag{2.34}$$

In view of (2.31), which implies (assuming f is integrable) $\hat{u}(0,\xi) = \hat{f}(\xi)$,

$$c(\xi) = \hat{f}(\xi). \tag{2.35}$$

The expression (2.34) of $\hat{u}(t, \xi)$ as a product of Fourier transforms leads to the *convolution* formula

$$u(t, x) = f * \varphi_t,$$

where $\hat{\varphi}_t(\xi) = e^{-tD|\xi|^2}$, i.e., φ is the *Gauss kernel*

$$\varphi_t(x) = (\frac{1}{\sqrt{4\pi Dt}})^k e^{-\frac{|x|^2}{4Dt}}, \tag{2.36}$$

and, therefore,

$$u(t, x) = \int_{R^k} f(y)\varphi_t(x - y)dy = (\frac{1}{\sqrt{4\pi Dt}})^k \int_{\mathbb{R}^k} f(y)e^{-\frac{|y-x|^2}{4Dt}}dy, \tag{2.37}$$

which is *linear* in f and may be expressed as

$$u(t, \cdot) = T_t f (= f * \varphi_t) \tag{2.38}$$

and satisfies

$$T_{t+s} f = T_t T_s f, \tag{2.39}$$

since $\varphi_{t+s} = \varphi_t * \varphi_s$. The operation (2.37) can be extended to the Banach space $\mathbb{B}(\mathbb{R}^k)$ or to its subspace $C_0(\mathbb{R}^k)$ of all real bounded continuous functions on $\mathbb{R}^k$ that *vanish at infinity* in the sense that given $\varepsilon > 0$, there is a compact set $K_\varepsilon \subset \mathbb{R}^k$, such that $|f(x)| < \varepsilon$ if $x \in K_\varepsilon^c$ (see Exercise 1).

Remark 2.3 Since continuous functions that vanish at infinity are also uniformly continuous (Exercise 3), the Banach space $C_0(\mathbb{R}^k)$ is sometimes a more convenient choice over $C_b(\mathbb{R}^k)$.

The semigroup resulting from this example is variously referred to as a *Gauss, normal, or heat semigroup*.

Note that $\varphi_t(y - x)dy = p(t; x, dy)$ is a transition probability (namely, that of a Brownian motion with diffusion coefficient $2D$). Thus, Example 4 belongs to the class of Markov semigroups of Example 1. Also, when the Equation (2.30) is expressed in the form (2.33), the solution (in the Fourier *frequency domain*) is of the form

$$\hat{u}(t, \xi) = e^{-tD|\xi|^2} \hat{f}(\xi), \tag{2.40}$$

i.e., $\widehat{T_t f}$ is a *multiplication* of $\hat{f}$ by $e^{-tD|\xi|^2}$. In this sense, the semigroup T_t may be viewed to be of the form $e^{tD\Delta}$ in the sense that $e^{tD\Delta} f$ is, in the frequency domain, $e^{-tD|\xi|^2}\hat{f}(\xi) = e^{t\widehat{A}(\xi)}\hat{f}(\xi)$, where $\widehat{A}$ is the representation of $D\Delta$ in the frequency domain (namely, multiplication by $-D|\xi|^2$). So, Example 4 also belongs to the class of Example 2 extended to unbounded A! A general method of extension is forthcoming in the form of the Hille–Yosida theorem.

Remark 2.4 Although (2.30) is the typical form of the heat equation in pde theory, in probability there is typically a factor of $\frac{1}{2}$ in front of $D\Delta$ on the right-hand side. As a result, the 4 in the formula (2.36) is replaced by a 2. This is natural from the perspective of standardizing the dispersion rate as a variance parameter noted above.

Let A be a linear operator defined on a linear subspace D_A (into $\mathscr{X}$). The notion of a *closed linear operator* is as follows.

Definition 2.2 A linear operator A having domain $\mathcal{D}_A$ in the Banach space $\mathscr{X}$ is said to be *closed* if its *graph* $G_A = \{(f, Af) : f \in D_A\}$ is a closed subset of $\mathscr{X} \times \mathscr{X}$ in the product topology.

Definition 2.3 Given a linear operator A defined on a linear space $D_A \subset \mathscr{X}$, denote the operator $(\lambda I - A) : \mathcal{D}_A \to \mathscr{X}$ simply by $\lambda - A$. The *resolvent set* of A is then

$$\rho(A) = \{\lambda \in F : \lambda - A \text{ is one-to-one and onto, } (\lambda - A)^{-1} \text{is bounded}\},$$

where F is the scalar field (real $\mathbb{R}$ or $\mathbb{C}$).

Proposition 2.3 *If* $\rho(A) \neq \emptyset$, *then* A *is closed.*

Proof Suppose $\lambda \in \rho(A)$; then, $(\lambda - A)^{-1}$ is bounded on $\mathscr{X}$ (into $\mathscr{X}$) and, therefore, obviously closed. Hence, $\lambda - A$ is closed, which implies A is closed. ■

Definition 2.4 For $\lambda \in \rho(A)$, $R_\lambda \equiv (\lambda - A)^{-1}$ is called the *resolvent operator*.

Proposition 2.4 (The Resolvent Identity) *Let* $\lambda, \mu \in \rho(A)$. *Then,*

$$(\lambda - A)^{-1} - (\mu - A)^{-1} \equiv R_\lambda - R_\mu = (\mu - \lambda)R_\lambda R_\mu. \tag{2.41}$$

In particular, R_λ *and* R_μ *commute.*

Proof Let $g \in \mathscr{X}$. There exists $f \in D_A$ such that $(\mu - A)f = g$. Hence,

$$\begin{aligned}
&((\lambda - A)^{-1} - (\mu - A)^{-1})g \\
&\quad = [(\lambda - A)^{-1} - (\mu - A)^{-1}](\mu - A)f
\end{aligned}$$

$$
\begin{aligned}
&= (\lambda - A)^{-1}(\mu - \lambda + \lambda - A)f - (\mu - A)^{-1}(\mu - A)f \\
&= (\mu - \lambda)R_\lambda f + f - f \\
&= (\mu - \lambda)R_\lambda R_\mu g.
\end{aligned}
$$

■

Definition 2.5 A *one-parameter contraction semigroup on* $\mathscr{X}$ is a family of maps $\{T_t : t > 0\}$ on $\mathscr{X}$ such that (i)–(iii) of the following hold.

(i) T_t is a bounded linear operator on $\mathscr{X}$, for each $t > 0$,
(ii) $T_{t+s} = T_t T_s$ for all $s, t > 0$,
(iii) $||T_t|| \leq 1$ for all $t > 0$.

The subspace $\mathscr{X}_0 = \{f \in \mathscr{X} : ||T_t f - f|| \to 0 \text{ as } t \downarrow 0\}$ is called the *center* of the semigroup.

Proposition 2.5 *The center $\mathscr{X}_0$ is a Banach space.*

Proof The main thing to check is that $\mathscr{X}_0$ is closed, as the vector space properties are easily seen to be inherited. Let $f_0 \notin \mathscr{X}_0$. Then, there exists $\delta > 0$ such that $||T_{t_n} f - f|| > \delta$ for a sequence $t_n \downarrow 0$. If $||f - f_0|| < \delta/4$, then

$$
||T_{t_n} f - f|| \geq ||T_{t_n} f_0 - f_0|| - ||T_{t_n}(f - f_0)|| - ||f - f_0|| > \delta - \frac{\delta}{4} - \frac{\delta}{4} = \frac{\delta}{2}.
$$

Hence, $f \notin \mathscr{X}_0$, proving closedness of $\mathscr{X}_0$. Thus, $\mathscr{X}_0$ is a Banach space. ■

Definition 2.6 In case $\mathscr{X} = \mathscr{X}_0$, the semigroup is said to be *of class* C_0 or *strongly continuous* on $\mathscr{X}$.

The reason for the latter terminology is the following (also see Exercise 1): On $[0, \infty)$, one has (take $T_0 f = f$)

$$
t \to T_t f \text{ is continuous for all } f \in \mathscr{X}_0. \tag{2.42}
$$

For one has

$$
||T_{t+s} f - T_t f|| = ||T_t(T_s f - f)|| \leq ||T_s f - f|| \to 0,
$$

as $s \downarrow 0$. Although we have made use of contraction in the last inequality, all that is needed is that $||T_t|| < \infty$, which is of course true. It should also be emphasized that

Proposition 2.6 *$\{T_t : t > 0\}$ is a strongly continuous semigroup on $\mathscr{X}_0$.*

Proof All that needs to be checked is that the semigroup leaves $\mathscr{X}_0$ *invariant*. That is, $f \in \mathscr{X}_0 \implies T_t f \in \mathscr{X}_0$. But this is proved by (2.42) and the display following it. ■

Definition 2.7 The *infinitesimal generator* A of a one-parameter semigroup $\{T_t : t > 0\}$ is defined by

$$Af = \lim_{t \downarrow 0} \frac{T_t f - f}{t} \tag{2.43}$$

for all $f \in \mathscr{X}$, for which this limit exists. The set of all such f is called the *domain of the infinitesimal generator* and denoted by $\mathcal{D}_A$. To include its domain, the pair $(A, \mathcal{D}_A)$ is often referred to as the infinitesimal generator. Clearly $\mathcal{D}_A \subset \mathscr{X}_0$.

Theorem 2.7 *Let $\{T_t : t > 0\}$ be a strongly continuous contraction semigroup on $\mathscr{X}$. Then,*

a. *$t \to T_t f$ is continuous on $[0, \infty)$ for all $f \in \mathscr{X}$.*
b. *$\overline{\mathcal{D}_A} = \mathscr{X}$ (i.e., A is densely defined).*
c. *$\mathcal{D}_A$ coincides with the range of R_λ for each $\lambda > 0$.*
d. *$T_t f \in \mathcal{D}_A$ for all $f \in \mathcal{D}_A$, T_t and A commute, i.e.,*

$$AT_t f = T_t A f. \tag{2.44}$$

e. *If $f \in \mathcal{D}_A$,*

$$T_t f - f = \int_0^t T_s A f ds. \tag{2.45}$$

f. *A is a closed linear operator.*
g. *If $Re\lambda > 0$, then $\lambda \in \rho(A)$, and*

$$||R_\lambda|| \le \frac{1}{Re\lambda}, \tag{2.46}$$

$$R_\lambda f = \int_0^\infty e^{-\lambda t} T_t f dt \quad \textit{for all } f \in \mathscr{X}. \tag{2.47}$$

Proof Part (**a**) has been already proved (see proof of (2.42)). Since $t \to ||T_t f - f||$ is bounded and continuous and $\lambda e^{-\lambda t} dt$ converges weakly to the Dirac measure δ_0 as $\lambda \to \infty$ (λ real), it follows that for all $f \in \mathscr{X}$, one has

$$\begin{aligned} ||\lambda \int_0^\infty e^{-\lambda t} T_t f dt - f|| &= ||\int_0^\infty (T_t f - f) \lambda e^{-\lambda t} dt|| \\ &\le \int_0^\infty ||T_t f - f|| \lambda e^{-\lambda t} dt \\ &\to \int_0^\infty ||T_t f - f|| \delta_0(dt) = ||f - f|| = 0. \end{aligned} \tag{2.48}$$

(Note that $T_t \to I$ strongly as $t \downarrow 0$, and one sets $T_0 = I$). Next, write

$$S_\lambda f = \int_0^\infty e^{-\lambda t} T_t f dt \quad (f \in \mathscr{X})$$

and note that

$$\begin{aligned}\frac{T_h S_\lambda f - S_\lambda f}{h} &= \frac{1}{h}\{\int_0^\infty e^{-\lambda t} T_{t+h} f dt - \int_0^\infty e^{-\lambda t} T_t f dt\} \\ &= \frac{1}{h}\{\int_h^\infty e^{-\lambda(t-h)} T_t f dt - \int_0^\infty e^{-\lambda t} T_t f dt\} \\ &= \frac{1}{h}\{(e^{\lambda h} - 1)\int_0^\infty e^{-\lambda t} T_t f dt - e^{\lambda h}\int_0^h T_t f dt\} \\ &\to \lambda S_\lambda f - f \text{ strongly as } h \downarrow 0.\end{aligned}$$

In other words, $S_\lambda f \in \mathcal{D}_A$ for all $f \in \mathscr{X}$ and all $\lambda > 0$ (or, in the complex cases, all λ such that $Re\lambda > 0$). Moreover,

$$A S_\lambda f = \lambda S_\lambda f - f,$$

or

$$(\lambda - A) S_\lambda f = f \quad \text{for all } f \in \mathscr{X} \tag{2.49}$$

Hence, $S_\lambda = (\lambda - A)^{-1}$, which exists and is bounded: $||S_\lambda|| \leq \frac{1}{\lambda}$. Thus, $R_\lambda = S_\lambda$ and

$$R_\lambda f = (\lambda - A)^{-1} f = \int_0^\infty e^{-\lambda t} T_t f dt, \quad (f \in \mathscr{X}), \tag{2.50}$$

proving (2.47). By (2.48),

$$\lim_{\lambda \to \infty} \lambda R_\lambda f = f \quad (\text{ for all } f \in \mathscr{X}). \tag{2.51}$$

Since the range of $\lambda R_\lambda \equiv \lambda(\lambda - A)^{-1}$ is the domain of $\lambda(\lambda - A)$ and, therefore, of A, A is closed, proving (f). Now, $\lambda R_\lambda f \in \mathcal{D}_A$ for all $f \in \mathscr{X}$. Hence, the left side of (2.51) implies $\overline{D}_A = \mathscr{X}$, proving (**b**) and (**c**). If $\mathscr{X}$ is a complex Banach space, one may similarly obtain

$$||R_\lambda f|| \leq ||f||/\mathrm{Re}\lambda \quad (\lambda \text{ such that } \mathrm{Re}\lambda > 0),$$

proving (**g**). Suppose $f \in \mathcal{D}_A$, $t > 0$. Then,

$$\frac{T_h T_t f - T_t f}{h} = \frac{T_t(T_h f - f)}{h} \to T_t A f \quad \text{as } h \downarrow 0,$$

proving that (**d**) holds, i.e., $T_t f \in \mathcal{D}_A$ and (2.44), Finally, (**e**) follows from the fundamental theorem of calculus, just as in the case of real-valued functions. ■

It is sufficient for our purposes to assume $\mathscr{X}$ to be a real Banach space. The following is the most important theorem of this chapter.

Theorem 2.8 (Hille–Yosida Theorem) *A linear operator A on $\mathscr{X}$ is the infinitesimal generator of a strongly continuous contraction semigroup if and only if*

a. *A is densely defined.*
b. *For all $\lambda > 0$, one has $\lambda \in \rho(A)$ and $||R_\lambda|| \leq 1/\lambda$.*

The proof of this theorem is deferred to Appendix A.

Theorem 2.9 (The Abstract Cauchy Problem) *Let $(A, \mathcal{D}_A)$ be the infinitesimal generator of a contraction semigroup $\{T_t : t \geq 0\}$ on $\mathscr{X}$. If $f \in \mathcal{D}_A$, then $u(t) = T_t f$ is the unique (strongly) differentiable solution to the equation*

$$\frac{du(t)}{dt} = Au(t) \quad (0 \leq t < \infty), \tag{2.52}$$

subject to the "initial condition"

$$||u(t) - f|| \to 0 \text{ as } t \downarrow 0. \tag{2.53}$$

Proof That $T_t f$ is a solution follows from Theorem 2.7, specifically (2.44), (2.45). To prove uniqueness, let $v(t)$ be any solution. Then, $s \to T_{t-s}v(s)$ is (strongly) differentiable for $0 \leq s \leq t$. Indeed, by (2.45),

$$\begin{aligned}
&\frac{d}{ds}(T_{t-s}v(s)) \\
&\quad = \lim_{h\to 0} \frac{T_{t-s-h}v(s+h) - T_{t-s}v(s)}{h} \\
&\quad = \lim_{h\to 0}\{\frac{T_{t-s-h}v(s+h) - T_{t-s}v(s+h)}{h} + \frac{T_{t-s}v(s+h) - T_{t-s}v(s)}{h}\} \\
&\quad = \lim_{h\to 0}\{-\frac{1}{h}\int_{t-s-h}^{t-s} T_u Av(s+h)du\} + T_{t-s}\frac{dv(s)}{ds} \\
&\quad = -T_{t-s}Av(s) + T_{t-s}Av(s) \\
&\quad = 0,
\end{aligned}$$

since $v(s+h) \in \mathcal{D}_A$ and $v(s)$ satisfies (2.52). Hence, $s \to T_{t-s}v(s)$ is independent of s. Set $s = 0, t$ to get $T_t v(0) = T_0 v(t) = v(t)$, i.e., $T_t f = v(t)$. ■

Example 5 (Birth–Death with Two Reflecting Barriers) $S = \{1, 2, \ldots, n\}$, $q_{ii} = -\lambda_i$ $(1 \le i \le n)$, $q_{i,i+1} = \lambda_i \mu_i$ $(1 \le i \le n-1)$, $q_{i,i-1} = \lambda_i \delta_i$ $(2 \le i \le n)$, $q_{ij} = 0$ for all other pairs (i, j). Here, $\mu_i > 0$ and $\delta_i > 0$ for all $2 \le i \le n-1$, $\mu_1 = 1$, $\delta_n = 1$, $\mu_i + \delta_i = 1$, $\lambda_i > 0$ for all i.

Example 6 (Pure Birth) $S = \{0, 1, 2, \ldots\}$, $q_{ii} = -\lambda_i$, $q_{i,i+1} = \lambda_i$ for all i, where $\lambda_i > 0$ for all i. If λ_i is a bounded sequence, then there exists a unique Markov semigroup with infinitesimal parameters $A = ((q_{ij}))$. An especially simple case is when $\lambda_i = \lambda$ for all i. In this case,

$$e^{tA} = e^{-\lambda t} e^{t\lambda B}, \qquad B = \begin{bmatrix} 0 & +1 & 0 & 0 & 0 & \ldots \\ 0 & 0 & +1 & 0 & 0 & \ldots \\ 0 & 0 & 0 & +1 & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{2.54}$$

Note that $(Bf)(i) = f(i+1)$, so that $(B^r f)(i) = f(i+r)$, and

$$\begin{aligned} p(t; i, j) &= (e^{tA})_{i,j} = e^{-\lambda t}(e^{t\lambda B} f_j)_i \ \left[f_j(i) := \delta_{ji}\right] \\ &= e^{-\lambda t} \sum_{r=0}^{\infty} \frac{(\lambda t)^r}{r!} (B^r f_j)(i) = e^{-\lambda t} \sum_{r=0}^{\infty} \frac{(\lambda t)^r}{r!} f_j(i+r) \\ &= \begin{cases} 0 & \text{if } j < i \\ e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!} & \text{if } j \ge i \end{cases} \end{aligned} \tag{2.55}$$

This provides an alternative approach in defining the homogeneous Poisson process.

Definition 2.8 The semigroup e^{tA} defined by (2.54) is referred to as the Poisson semigroup, and the Markov process is the homogeneous Poisson process, with intensity parameter λ.

Note: To accommodate the possibility that a Markov chain may leave the state space in finite time t, it is fruitful to allow *substochastic* transition probabilities: $\int_S p(t; x, dy) < 1$, and therefore, in Definition (2.1), replace (i) by

$$(i') : B \to p(t; x, B) \textit{ is a measure on } S \textit{ such that } p(t; x, S) \le 1.$$

We permit the substochastic relaxation for the transition probabilities without further remark. In contrast, transition probabilities for which $\int_S p(t; x, dy) = 1$ for all $x \in S$, $t \ge 0$, are said to be *stochastic* or *conservative*.

If one has $p(t; x, S) < 1$ for some $t > 0$, $x \in S$, one may introduce a fictitious state s_∞ and consider the new state space $\overline{S} = S \cup \{s_\infty\}$, with the σ-field $\overline{\mathcal{S}} = \mathcal{S} \cup \{B \cup \{s_\infty\} : B \in \mathcal{S}\}$, and transition probability

$$\begin{aligned}
\overline{p}(t; x, B) &= p(t; x, B) && \text{if } x \in S,\ B \in \mathcal{S}, \\
\overline{p}(t; x, \{s_\infty\}) &= 1 - p(t; x, S) && \text{if } x \in S, \\
\overline{p}(t; s_\infty, \{s_\infty\}) &= 1 && \text{for all } t \geq 0.
\end{aligned} \tag{2.56}$$

It is then straightforward to check conditions (i), (ii), and (iii) of Definition 2.1. There are other ways of augmenting p also; some of these will be discussed later in this chapter.

Example 7 (Convolution Semigroups) Let $S = \mathbb{R}^k$ ($k \geq 1$). Suppose $\{Q_t : 0 \leq t < \infty\}$ is a one-parameter family of probability measures on $\mathbb{R}^k$, i.e., on its Borel σ-field, such that

$$Q_t * Q_s = Q_{t+s}, \quad \lim_{t \downarrow 0} Q_t = \delta_0, \tag{2.57}$$

where the limit is in the sense of weak convergence of probability measures:

$$\int f dQ_t \to \int f d\delta_0 = f(0) \quad ast \downarrow 0,$$

for all real, bounded continuous f on $\mathbb{R}$. In a convenient abuse of notation, one may define linear operators

$$Q_t f(x) = \int_{\mathbb{R}^k} f(x+y) Q_t(dy) = \int_{\mathbb{R}^k} f(y') p(t; x, dy'), \quad f \in C_b(\mathbb{R}^k : \mathbb{R}^k), \tag{2.58}$$

where as a Markov semigroup, writing $B - x = \{y - x : y \in B\}$, $B \subset \mathbb{R}^k$, the transition probabilities (see (2.2)) are given by

$$p(t; x, B) = Q_t(B - x), \quad B \in \mathcal{B}(\mathbb{R}^k), x \in \mathbb{R}^k, t \geq 0. \tag{2.59}$$

Then,

$$\begin{aligned}
Q_{t+s} f(x) &= \int_{\mathbb{R}^k} f(x+y) Q_{t+s}(dy) \\
&= \int_{\mathbb{R}^k} f(x+y)(Q_t * Q_s)(dy) \\
&= \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} f(x+y_1+y_2) Q_t(dy_1) Q_s(dy_2)
\end{aligned}$$

$$= \int_{\mathbb{R}^k} (\int_{\mathbb{R}^k} f(x + y_1 + y_2) Q_s(dy_2)) Q_t(dy_1)$$

$$= \int_{\mathbb{R}^k} (Q_s f)(x + y_1) Q_t(dy_1) = (Q_t Q_s f)(x) \tag{2.60}$$

Therefore, $\{Q_t : 0 \leq t < \infty\}$ is a one-parameter semigroup on $C_b(\mathbb{R}^k)$. It is called a *convolution semigroup*. Since, by the weak convergence (2.57)

$$Q_t f(x) - f(x) = \int_{\mathbb{R}^k} [f(x + y) - f(x)] Q_t(dy)$$

goes to zero uniformly in x, if f is uniformly continuous (and bounded), $\{Q_t : 0 \leq t < \infty\}$ is *strongly continuous* on $C_0(\mathbb{R}^k)$, the space of continuous functions vanishing at infinity.

Convolution semigroups have already occurred in previous examples:

1. $Q_t(dy) = \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{y^2}{2\sigma^2 t}} dy$, (see Remark 2.4).
2. $Q_t(\{nh\}) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$, $\quad Q_t(\mathbb{R} \setminus \{nh : n = 0, 1, 2, \ldots\}) = 0$, where h is a given nonzero number.

Note that if $Q_t^{(1)}, Q_t^{(2)}, \ldots, Q_t^{(m)}$ are m convolution semigroups, then the semigroup corresponding to $Q_t^{(1)} * Q_t^{(2)} * \ldots * Q_t^{(m)} = Q_t$ is another, i.e.,

$$(Q_t^{(1)} * Q_t^{(2)} * \ldots * Q_t^{(m)}) * (Q_s^{(1)} * Q_s^{(2)} * \ldots * Q_s^{(m)})$$

$$= (Q_t^{(1)} * Q_s^{(1)}) * (Q_t^{(2)} * Q_s^{(2)}) * \ldots (Q_t^{(m)} * Q_s^{(m)})$$

$$= Q_{t+s}^{(1)} * Q_{t+s}^{(2)} * \ldots * Q_{t+s}^{(m)}. \tag{2.61}$$

It turns out that all convolution semigroups may be obtained from the Gauss (or normal) and Poisson semigroups as components and using (2.61) and taking limits. This is the topic of Chapter 5.

Before concluding this chapter, let us briefly record the abstract forms of Kolmogorov's *backward* and *forward equations*. The abstract equations are (see (2.44))

$$\frac{d}{dt} T_t f = A T_t f \text{ (backward equation)}, \qquad \frac{d}{dt} T_t f = T_t A f \text{ (forward equation)}. \tag{2.62}$$

For Markov chains on countable state spaces S in Example 3, these take the form

$$\sum_{j\in S} f(j)p'(t;i,j) = \sum_{j\in S} q_{ij} \sum_{k\in S} f(k)p(t;j,k),\ i \in S,$$
$$\sum_{j\in S} f(j)p'(t;i,j) = \sum_{j\in S} (\sum_{k\in S} q_{jk} f(k))p(t;i,j),\ i \in S. \tag{2.63}$$

Much will be discovered about these equations for important classes of Markov processes in forthcoming chapters.

Exercises

1. Consider the k-dimensional Gauss semigroup $\{T_t\}$ defined by the transition probabilities $p(t;x,dy) = \frac{1}{(2\pi t)^{\frac{k}{2}}} e^{-\frac{1}{2t}|y-x|^2} dy$ in Example 4 with $\mathscr{X} = \mathbb{B}(\mathbb{R}^k)$, with the sup-norm. Show that the center is the space of continuous functions on $\mathbb{R}^k$ that vanish at infinity. In particular, show that $\{T_t\}$ is not strongly continuous as a semigroup on the space of bounded continuous functions.
2. In Example 4, let $\mathscr{X} = L^2(\mathbb{R}^k)$, with the L^2-norm: $||f|| = \sqrt{\int_{\mathbb{R}^k} |f|^2 d\mu}$, μ being Lebesgue measure on $\mathbb{R}^k$. (a) Show that T_t defined on $\mathscr{X}$ is a contraction semigroup, with generator Δ. (b) Is this semigroup strongly continuous?
3. (a) Show that the space of continuous functions on $\mathbb{R}^k$ that vanish at infinity is a separable Banach space. (b) Show that a continuous function on $\mathbb{R}^k$ that vanishes at infinity is uniformly continuous.
4. *(Positive Maximum Principle)* Consider a strongly continuous Markov semigroup on the Banach space $\mathscr{X} \equiv C_0(S)$ of continuous functions vanishing at infinity on a locally compact metric space S, endowed with the sup-norm, having infinitesimal generator $(A, \mathcal{D}_A)$. Suppose that $f \in \mathcal{D}_A$, and $0 \le f(x_0) = \sup_{x\in S} f(x)$ for some $x_0 \in S$. Show that $Af(x_0) \le 0$. [*Hint*: Note that $T_t f(x_0) - f(x_0) = \int_S f(y)(p(t;x_0,dy) - f(x_0) \le f(x_0)(p(t;x_0,S)-1) \le 0$, and use the definition of $Af(x_0)$.]
5. Let $(S, \mathcal{S})$ be the state space of a continuous parameter Markov process $X = \{X_t : t \ge 0\}$. Assume that S is a Polish space and $\mathcal{S}$ is its Borel σ-field. Prove that the transition probabilities of the Markov process satisfy (i), (ii), and (iii) of Definition 2.1. [*Hint*: On a Polish space, one may assume the existence of regular conditional probabilities; see BCPT pp. 29–30, Breiman 1968, p78.]
6. *(Compound Poisson Process)* Let $N = \{N_t : t \ge 0\}$ be a homogeneous Poisson process with intensity $\rho > 0$. Let $Y_1, Y_2, \ldots$ be an iid sequence of integer-valued random variables, independent of $\{N_t : t \ge 0\}$. Define an integer-valued

compound Poisson process by $X_t = \sum_{j=0}^{N_t} Y_j, \quad t \geq 0$, where $N_0 = 0, Y_0 = x_0 \in \mathbb{Z}$. (i) Prove that $X_t, t \geq 0$, is a process with independent increments and is, therefore, Markov. (ii) Find the infinitesimal generator Q of the Markov process $\{X_t : t \geq 0\}$ in (i) in terms of the probability mass function f of Y_1, and the Poisson parameter, $\rho > 0$.

7. *(Continuous Time Simple Symmetric Random Walk)* Let the probability generating function f in Exercise 6 be that of symmetric Bernoulli, $f(1) = f(-1) = \frac{1}{2}$. (i) Compute the infinitesimal generator Q. (ii) Show that $p_{ii}(t) = 1 - \rho t + \frac{3}{2}\rho^2 t^2 + O(t^3)$, as $t \downarrow 0$. [*Hint*: $p_{ii}(t) = \sum_{m=0}^{\infty} q_{ii}^{(m)} \frac{t^m}{m!}$, where $Q^m = ((q_{ij}^{(m)}))$.]
8. *(Deterministic Motion)* Consider the initial value problem $\frac{\partial u(t,x)}{\partial t} = \mu \cdot \nabla u$, where $\mu = (\mu_1, \ldots, \mu_k) \in \mathbb{R}^k$, $\nabla = \sum_{j=1}^{k} \frac{\partial}{\partial x^j}$, $(t > 0, x \in \mathbb{R}^k)$, with initial condition $u(0, x) = f(x)$, $f \in L^1(\mathbb{R}^k)$. Apply the Fourier method of Example 4, to interpret the semigroup e^{tA} for an unbounded operator A in this case. [*Hint*: This involves the Fourier transform of the point mass measure $\delta_a(dx)$ (see BCPT, p.82).]
9. *(Resolvent of Brownian Motion)* Show that the resolvent of the one-dimensional standard Brownian motion semigroup can be expressed as $R_\lambda f(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\sqrt{2\lambda}|y-x|} f(y)dy$.
10. *(Resolvent of the Poisson Process)* Calculate the resolvent of the Poisson semigroup with intensity parameter $\rho > 0$. [*Hint*: Use a change of variables to express the integral in terms of the Gamma function $\Gamma(s) = \int_0^\infty t^{s-1} e^{-t} dt$, noting that $\Gamma(k + 1) = k!$ for nonnegative integers k.]
11. Consider an integer-valued continuous parameter Markov process. Show for bounded f that $R_\lambda f(i) = \sum_j \hat{p}_{ij}(\lambda) f(j)$, where, for $\lambda > 0$, $\hat{p}_{ij}(\lambda) = \int_0^\infty e^{-\lambda t} p_{ij}(t) dt$.
12. *(Homogeneous Cox Process)* The Cox process, or doubly stochastic Poisson process, $X = \{X_t : t \geq 0\}$ is defined for a given nonnegative random variable Λ with distribution μ as conditionally given ρ, a Poisson process with intensity Λ. (i) In the case $\mathbb{E}\Lambda^2 < \infty$, show that $\mathrm{Var}(X_t) = \mathbb{E}\Lambda t + \mathrm{Var}(\Lambda) t^2$. (ii) Use (i) to conclude that X cannot have independent increments if Λ is *not* a.s. constant. [*Hint*: Note homogeneity of increments and consider what that would imply for independent increments.] (iii) Show that X is not Markov unless Λ is a.s. constant. [*Hint*: Check the *failure* of the following special case of the semigroup property $p_{00}(t+s) = p_{00}(t) p_{00}(s)$, $s, t > 0$, for this process, unless Λ is an a.s. constant. Assume the semigroup property and take $s = t$. Express the consequence in terms of $\mathrm{Var}(e^{-\Lambda t})$.]

Chapter 3
Regularity of Markov Process Sample Paths

This chapter provides general constructions of Markov processes on compact and locally compact separable state spaces from their Feller transition probabilities and arbitrary initial distributions.

In this chapter, Markov processes, having a locally compact, separable state space, are constructed with right-continuous sample paths having left limits, often referred to as *càdlàg paths*. An alternative construction is presented in Chapter 21 without topological restrictions on the state space beyond that of being a metric space; see Theorems 21.2 and 21.4.

Consider the Poisson process $\{X_t : t \geq 0\}$ with intensity parameter $\lambda > 0$ in Definition 1.3 from Chapter 1. Although the sample paths are step functions with unit jumps, for any $\varepsilon > 0$, and fixed t, one has a weak form of "continuity" in that

$$\lim_{s \to t} P(|X_t - X_s| > \varepsilon) = 1 - e^{-\lambda|t-s|} \to 0 \text{ as } s \to t.$$

With this example in mind, the following is a natural definition.

Definition 3.1 A stochastic process $\{X_t\}$ with values in a metric space (S, d) is said to be *continuous in probability*, or *stochastically continuous*, at t if for each $\varepsilon > 0$,

$$\lim_{s \to t} P(d(X_t, X_s) > \varepsilon) = 0.$$

If this holds for all t, then the process is said to be continuous in probability, or stochastically continuous.

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_3

Definition 3.2 Stochastic processes $\{\tilde{X}_t\}$ and $\{X_t\}$ defined on a common probability space $(\Omega, \mathcal{F}, P)$ are said to be *stochastically equivalent* if $P(X_t = \tilde{X}_t) = 1$ for each $t \geq 0$. In this case, $\{\tilde{X}_t\}$ is also called a *version* of $\{X_t\}$.

Theorem 3.1 *Let $\{X_t\}$ be a submartingale or supermartingale with respect to a filtration $\{\mathcal{F}_t\}$. Suppose that $\{X_t\}$ is continuous in probability at each $t \geq 0$. Then, there is a stochastic process $\{\tilde{X}_t\}$ stochastically equivalent to $\{X_t\}$, such that with probability one, the sample paths of $\{\tilde{X}_t\}$ are bounded on compact intervals $a \leq t \leq b$, $(a, b \geq 0)$, and are right-continuous and have left-hand limits at each $t > 0$, i.e., having càdlàg sample paths.*

Proof Fix a rational $T > 0$, and let Q_T denote the set of rational numbers in $[0, T]$. Write $Q_T = \bigcup_{n=1}^{\infty} R_n$, where each R_n is a *finite* subset of $[0, T]$ and $T \in R_1 \subset R_2 \subset \cdots$. By Doob's maximal inequality (see Chapter 1), one has

$$P(\max_{t \in R_n} |X_t| > \lambda) \leq \frac{\mathbb{E}|X_T|}{\lambda}, \qquad n = 1, 2, \ldots.$$

Therefore,

$$P(\sup_{t \in Q_T} |X_t| > \lambda) \leq \lim_{n\to\infty} P(\max_{t \in R_n} |X_t| > \lambda) \leq \frac{\mathbb{E}|X_T|}{\lambda}.$$

In particular, the paths of $\{X_t : t \in Q_T\}$ are *bounded* with probability one. Let (c, d) be any interval in $\mathbb{R}$ and let $U^{(T)}(c, d)$ denote the number of upcrossings of (c, d) by the process $\{X_t : t \in Q_T\}$. Then, $U^{(T)}(c, d)$ is the limit of the number $U^{(n)}(c, d)$ of upcrossings of (c, d) by $\{X_t : t \in R_n\}$ as $n \to \infty$. By the upcrossing inequality (see BCPT[1] Theorem 3.11), one has

$$\mathbb{E}U^{(n)}(c, d) \leq \frac{\mathbb{E}|X_T| + |c|}{d - c}.$$

Since the $U^{(n)}(c, d)$ are nondecreasing with n, it follows that $U^{(T)}(c, d)$ is a.s. finite. Taking unions over all (c, d), with c, d rational, it follows that with probability one, $\{X_t : t \in Q_T\}$ has only finitely many upcrossings of any finite interval. In particular, therefore, left- and right-hand limits must exist at each $t < T$ a.s. To construct a right-continuous version of $\{X_t\}$, define $\tilde{X}_t = \lim_{s\to t^+, s \in Q_T} X_s$ for $t < T$. That $\{\tilde{X}_t\}$ is in fact stochastically equivalent to $\{X_t\}$ now follows from continuity in probability; i.e., $\tilde{X}_t = p{-}\lim_{s\to t^+} X_s = X_t$ since a.s. limits and limits in probability must a.s. coincide. Since T is arbitrarily large, the proof is complete. ■

[1] Throughout, BCPT refers to Bhattacharya and Waymire (2016), A Basic Course in Probability Theory.

The above martingale regularity theorem can now be applied to the problem of constructing Markov processes with regular sample path properties. Let us recall :

Definition 3.3 Markov transition probabilities $p(t; x, dy)$ are said to have the *Feller property* if for each (bounded) continuous function f, the function $x \to \int_S f(y)p(t; x, dy) = \mathbb{E}_x f(X_t)$ is continuous. In this case, the corresponding Markov process is referred to as a Feller process.

Theorem 3.2 *Let $\{X_t : t \geq 0\}$ be a Feller Markov process with a locally compact and separable metric state space S with (Feller) transition probability function*

$$(i) \quad p(t; x, B) = P(X_{t+s} \in B \mid X_s = x), x \in S,\ s, t \geq 0,\ B \in \mathcal{B}(S),$$

such that for each $\varepsilon > 0$,

$$(ii) \quad p(t; x, B^c(x : \varepsilon)) = \mathrm{o}(1) \text{ as } t \to 0^+ \text{ uniformly for } x \in S.$$

Then, there is a stochastically equivalent version $\{\tilde{X}_t\}$ of $\{X_t\}$ a.s. having càdlàg sample paths at each $t < \zeta_\infty$, where

$$\zeta_\infty = \inf\{t > 0 : \overline{X}_t \notin S\},$$

is an extended real-valued random variable, referred to as the explosion time for the process $\{\overline{X}_t\}$, to be constructed in the proof.

Proof Consider first the case that S is a compact metric space. It is enough to prove that $\{X_t\}$ a.s. has left- and right-hand limits at each t, for then $\{X_t\}$ can be modified as $\tilde{X}_t = \lim_{s \to t^+} X_s$, which by stochastic continuity will provide a version of $\{X_t\}$. It is not hard to check stochastic continuity from (i) and (ii). Let $f \in C(S)$ be an arbitrary continuous function on S. Consider the semigroup $\{T_t\}$ acting on $C(S)$ by

$$T_t f(x) = \mathbb{E}_x(X_t), \quad x \in S,\ t \geq 0.$$

For $\lambda > 0$, denote its resolvent by

$$R_\lambda f(x) = \int_0^\infty e^{-\lambda s} T_s f(x)\, ds, \qquad x \in S.$$

Consider the process $\{Y_t\}$ defined by

$$Y_t = e^{-\lambda t} R_\lambda f(X_t), \qquad t \geq 0, \tag{3.1}$$

where $f \in C(S)$ is a fixed but arbitrary *nonnegative* function on S. Since $\mathbb{E}|Y_t| < \infty$, Y_t is $\mathcal{F}_t$-measurable, $T_t f(x) \geq 0$ ($x \in S, t \geq 0$), and

$$\mathbb{E}\{Y_{t+h}|\mathcal{F}_t\} := e^{-\lambda(t+h)}\mathbb{E}\{R_\lambda f(X_{t+h})|\mathcal{F}_t\} = e^{-\lambda(t+h)} T_h R_\lambda f(X_t)$$

$$= e^{-\lambda t} \int_0^\infty e^{-\lambda(s+h)} T_{s+h} f(X_t)\, ds = e^{-\lambda t} \int_h^\infty e^{-\lambda s} T_s f(X_t)\, ds$$
$$\leq e^{-\lambda t} \int_0^\infty e^{\lambda s} T_s f(X_t)\, ds = Y_t, \tag{3.2}$$

it follows that $\{Y_t\}$ is a nonnegative supermartingale with respect to $\mathcal{F}_t = \sigma\{X_s\ s \leq t\}$, $t \geq 0$. Applying the martingale regularity result of Theorem 3.1, we obtain a version $\{\tilde{Y}_t\}$ of $\{Y_t\}$, whose sample paths are a.s. right continuous with left-hand limits at each t. Thus, the same is true for $\{\lambda e^{\lambda t} Y_t\} = \{\lambda R_\lambda f(X_t)\}$. Since $\lambda R_\lambda f \to f$ uniformly as $\lambda \to \infty$, the process $\{f(X_t)\}$ must, therefore, a.s. have left- and right-hand limits at each t. The same will be true for any $f \in C(S)$, since one can write $f = f^+ - f^-$ with f^+ and f^- continuous *nonnegative* functions on S. So we have that for each $f \in C(S)$, $\{f(X_t)\}$ is a process whose left- and right-hand limits exist at each t (with probability one). As remarked at the outset, it will be enough to argue that this means that the process $\{X_t\}$ will a.s. have left- and right-hand limits. Since S is a compact metric space, it has a countable dense subset $\{x_n\}$. The functions $f_n : S \to \mathbb{R}$ defined by $f_n(x) = \rho(x_n, x)$, $x \in S$, are continuous for the metric ρ, and separate points of S in the sense that if $x \neq y$, then for some n, $f_n(x) \neq f_n(y)$. In view of the above, for each n, $\{f_n(X_t)\}$ is a process whose left- and right-hand limits exist at each t with probability one. From the countable union of events of probability 0 having probability 0, it follows that, with probability one, for all n the left- and right-hand limits exist at each t for $\{f_n(X_t)\}$. One may restrict the sample space to a set of sample points of probability one, such that $f(X_t)$ has left- and right-hand limits for every sample point, for every n. But this means that the left- and right-hand limits exist at each t for $\{X_t\}$, since the f_n's separate points; i.e., if either limit, say left, fails to exist at some t', then the sample path $t \to X_t$ must have at least two distinct limit points as $t \to t'^-$, contradicting the corresponding property for all the processes $\{f_n(X_t)\}, n = 1, 2, \ldots$. In this case, one has $\zeta_\infty = \infty$ *a.s.*, i.e., no explosion occurs since $\sup\{f_n(x) : x \in S, n \geq 1\} < \infty$.

In the case that S is locally compact, one may adjoin a point at infinity, denoted $s_\infty (\notin S)$, to S. The topology of the one point compactification on $\bar{S} = S \cup \{s_\infty\}$ defines a neighborhood system for s_∞ by complements of compact subsets of S. Let $\bar{\mathcal{B}}(S)$ be the σ-field generated by $\{\mathcal{B}, \{s_\infty\}\}$. The transition probability function $p(t; x, B)$, $(t \geq 0, x \in S, B \in \mathcal{B}(S))$ is extended to $\bar{p}(t; x, B)$ $(t \geq 0, x \in \bar{S}, B \in \bar{\mathcal{B}}(\bar{S}))$ by making s_∞ an absorbing state; i.e., $\bar{p}(t; s_\infty, B) = 1$ if and only if $s_\infty \in B$, $B \in \bar{\mathcal{B}}$, otherwise $\bar{p}(t; s_\infty, B) = 0$. If the conditions of Theorem 3.2 are fulfilled for $\bar{p}$, then one obtains a regular process $\{\bar{X}_t\}$ with state space $\bar{S}$. The basic Theorem 3.2 provides a process $\{\tilde{X}_t\ t < \zeta_\infty\}$ with state space S, whose sample paths are right-continuous with left-hand limits by restricting to $t < \zeta_\infty$. ■

Exercises

1. (i) Prove that two stochastically equivalent processes have the same finite-dimensional distributions.
   (ii) Prove that under the hypothesis of Theorem 3.2, $\{X_t\}$ is stochastically continuous.
   (iii) Prove that, under the hypothesis of Theorem 3.2, $\lambda R_\lambda f \to f$ uniformly as $\lambda \to \infty$. [*Hint*: Check that as $\lambda \to \infty$, $\lambda e^{-\lambda t} dt$ converges weakly to $\delta_0(dt)$. Also, under the condition (ii), the supnorm $||T_t f - f|| \to 0$ as $t \to 0$.]
2. Assume the framework of Theorem 3.2. Give conditions for $u(t,x) = \mathbb{E}_x u_0(X(t)),\ t \geq 0$, to solve the abstract Cauchy problem stated in Theorem 2.9. [*Hint*: Consider continuity and decay conditions on the initial data.]
3. Show that there is a càdlàg version of the Poisson process defined by the transition probabilities in (2.55).
4. Show that there is a càdlàg version of the process associated with the transition probabilities (2.59) defined by the convolution semigroup.
5. *(An Explosive Process)* Let $\{T_n\}_{n=1}^\infty$ be i.i.d. mean one exponentially distributed random variables, and $\{\lambda_n\}$, a sequence of positive numbers. Define $X_t = 0$ if $0 \leq t < \lambda_1^{-1} T_1$, and $X_t = n$ if $\sum_{j=1}^{n-1} \lambda_j^{-1} T_j \leq t < \sum_{j=1}^{n} \lambda_j^{-1} T_j$ for $n = 2, 3, \ldots$. Let $S = \{0, 1, 2, \ldots\}$.

   (i) Show that $\{X_t\}$ is a Markov process. [*Hint*: Use induction on n to show for fixed $0 < s < t$, $P(X_t \leq n | X_u, u \leq s) = P(X_t \leq n | X_s)$ on $[X_s = m]$ for each $m = 0, 1, \ldots, n$, noting that $\sigma(X_u, u \leq s) = \sigma(X_s, \lambda_j^{-1} T_j, j \leq X_s)$ and $[X_s = m] = [S_m \leq s < S_{m+1}]$, $S_m = \sum_{j=1}^m \lambda_j^{-1} T_j$.]
   (ii) Assume $\sum_{n=1}^\infty \frac{1}{\lambda_n} < \infty$. Show that $\zeta_\infty = \inf\{t > 0 : X_t \notin S\} = \sum_{n=1}^\infty \lambda_n^{-1} T_n < \infty$ with probability one. [*Hint*: Calculate $\mathbb{E} \sum_{n=1}^\infty \lambda_j^{-1} T_j$.]
   (iii) (*Yule Process*) The Yule process may be defined by the choice $\lambda_j = j\lambda,\ j = 1, 2 \ldots$ (compare Chapter 1, Exercise 3). Show, more generally, if $\sum_{j=1}^\infty \frac{1}{\lambda_j} = \infty$, then $\{X_t\}$ is nonexplosive. [*Hint*: Show that $\mathbb{E} \exp\{-\sum_{j=1}^\infty \frac{1}{\lambda_j} T_j\} = 0$.]

Chapter 4
Continuous Parameter Jump Markov Processes

Two problems are addressed in this chapter. The first is to compute the space-time structure of Markov processes on a general state space having piecewise constant sample paths. The second is to provide a construction of continuous parameter Markov processes having prescribed infinitesimal parameters, including a treatment of solutions to the associated Cauchy problem (backward differential equations). This includes Feller's method of successive approximations to analytically construct the minimal solution to both the backward and forward equations as possibly substochastic transition probabilities having prescribed infinitesimal parameters and satisfying the Chapman–Kolmogorov equations.

Let $\{X_t : t \geq 0\}$ be a Markov process on a measurable state space $(S, \mathcal{S})$, whose sample paths are piecewise constant between jumps. Such processes are said to be *pure jump Markov processes*.

We begin by considering the joint distribution of the time spent in the initial state and its displacement at the time of the first jump. This analysis is facilitated by a combination of the Markov property and the pure jump sample path regularity.

Unless otherwise stated, it will be assumed in this chapter that S is a locally compact and separable metric space with Borel σ-field $\mathcal{S}$. In particular, a countable set S with the discrete topology provides an important special case.

Theorem 4.1 *Assume that* $\{X_t : 0 \leq t < \infty\}$ *is a Markov process on a locally compact and separable metric space* $(S, \mathcal{S})$ *starting at* $X_0 = x$, *and let* $T_0 \equiv T_0(x) := \inf\{t > 0 : X_t \neq x\}$. *Then, there is a nonnegative measurable function* λ *on* S *such that* T_0 *has an exponential distribution with parameter* $\lambda(x)$.

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_4

Moreover, X_{T_0} and T_0 are independent random variables. In the case $\lambda(x) = 0$, the degeneracy of the exponential distribution can be interpreted to mean that x is an absorbing state, i.e., $P_x(T_0 = \infty) = 1$.

Proof Define $\psi(t, x, B) = P_x(T(x) > t, X_{T(x)} \in B)$, $x \in S$, $B \in \mathcal{S}$. Let X_s^+ denote the after-s process defined by

$$(X_s^+)_t = X_{s+t}, \quad t \geq 0.$$

Then, conditioning on $\mathcal{F}_s = \sigma(X_u : 0 \leq u \leq s)$, one has $[T(x) > s] \in \mathcal{F}_s$, so that

$$\begin{aligned}
\psi(s+t, x, B) &= P_x(T(x) > s, T(X_s^+) > t, X_{T(x)} \in B) \\
&= P_x(T(x) > s, T(X_s^+) > t, X_{T(X_s)} \in B) \\
&= \mathbb{E}_x(\mathbf{1}_{[T(x)>s]}\psi(t, X_s, B)) \\
&= \mathbb{E}_x(\mathbf{1}_{[T(x)>s]}\psi(t, x, B)),
\end{aligned}$$

since $X_s = X_0 = x$ on $[T(x) > s]$. Letting $t \downarrow 0$ in (4.1), since $\psi(t, x, B)) \uparrow P_x(X_{T(x)} \in B)$ and $\psi(s+t, x, B) \uparrow \psi(s, x, B)$, one has

$$P_x(T(x) > s, X_{T(x)} \in B) = P_x(T(x) > s)P_x(X_{T(x)} \in B), \tag{4.1}$$

proving independence. Next, observe that $\psi(t, x, S) = P_x(T(x) > t)$, and apply the last line of (4.1) to $B = S$, to get

$$P_x(T(x) > s+t) = P_x(T(x) > s)P_x(T(x) > t). \tag{4.2}$$

The only nonnegative, right continuous solutions to this equation are of the form

$$P_x(T_0(X) > t) = e^{-\lambda(x)t}, \quad t \geq 0,$$

for a constant $\lambda(x)$ (Exercise 1). Since $t \to P_x(T_0(X) > t)$ is nondecreasing, $\lambda(x) \geq 0$. The measurability of $x \to \lambda(x)$ follows from that of $x \to P_x(T_0(X) > t)$. ■

Remark 4.1 For the specification of the infinitesimal generator A of the Markov process in Theorem 4.1, let $\mathbb{B}(S)$ denote the Banach space of bounded measurable functions on S endowed with the sup norm, and let $T_t(t \geq 0)$ be the semigroup of transition operators on $\mathbb{B}(S)$. Then, for $f \in \mathbb{B}(S)$, denoting by K the transition operator of the Markov chain with transition probability $k(x, dy)$,

$$\begin{aligned}
T_t f(x) &= \exp\{-\lambda(x)t\}f(x) + \int_{(0,t]}\int_S (\lambda(x)\exp\{-\lambda(x)u\}f(y)k(x, dy)du \\
&= \exp\{-\lambda(x)t\}f(x) + (1 - \exp\{-\lambda(x)t\})Kf(x)).
\end{aligned} \tag{4.3}$$

For $f \in \mathcal{D}_A$, using

$$\exp\{-\lambda(x)t\} = 1 - \lambda(x)t = o(t),$$

$$Af(x) = \lim_{t\downarrow} \frac{T_t f(x) - f(x)}{t} = -\lambda(x)f(x) + \lambda(x)Kf(x) \ (f \in \mathcal{D}_A), \tag{4.4}$$

the domain $\mathcal{D}_A$ comprising all f for which the limit holds uniformly on S. In particular, if $\lambda(x)$ is bounded on S, then A is a bounded operator on $\mathbb{B}(S)$. If $\lambda(x)$ is continuous and $x \to k(x, dy)$ is weakly continuous, then it follows from the above that $x \to p(t; x, dy)$ is weakly continuous for all $t > 0$. However, as stated above, in the case $\lambda(x)$ is bounded, no weak continuity condition is required for the Markov process generated by A. Note that on countable state spaces S, which are the most common and important in many applications, the continuity properties are automatic under the discrete topology, since every function on S is continuous. Finally, if $\lambda(x)$ is unbounded, explosion may occur. Also, $k(x, dy)$ may be a subprobability kernel, i.e., $k(x, S)$ may be less than 1 for some x.

For the following corollary to Theorem 4.1, let

$$\begin{aligned} &\tau_0 \equiv \tau_0(X) := 0, && \tau_n \equiv \tau_n(X) := \inf\{t > \tau_{n-1} : X_t \neq X_{\tau_{n-1}}\}, \\ &T_0 \equiv T_0(X) := \tau_1, && T_n \equiv T_n(X) := \tau_{n+1} - \tau_n \quad n \geq 1. \end{aligned} \tag{4.5}$$

Corollary 4.2 *Let $\{X_t : t \geq 0\}$ be a continuous parameter Markov process on a locally compact and separable metric space having the infinitesimal parameters $\lambda(x), k(x, dy)$ and initial state x. Then, $\{Y_n := X_{\tau_n} : n = 0, 1, 2, \ldots\}$ is a discrete parameter Markov process having the transition probabilities $k(x, dy)$. Also, conditionally given $\{Y_n : n \geq 0\}$, the holding times $T_0(X)$, $T_1(X)$, $T_2(X)$, … are independent exponentially distributed random variables with parameters $\lambda(x_0)$, $\lambda(x_1)$, $\lambda(x_2)$, … on the event $[Y_0 = x_0, Y_1 = x_1, Y_2 = x_2, \ldots]$.*

Proof Theorem 4.1 implies that $T_0(X)$ and the after-$T_0(X)$ process, $X^+_{T_0}$, are independent by the strong Markov property (Exercise 2). ■

The discrete parameter Markov process $\{Y_n\}$ defined in Corollary 4.2 is typically referred to as the *skeleton process*.

Next, let us show that conversely given the infinitesimal parameters $\lambda(x), k(x, dy)$, there is a semigroup of, possibly substochastic, transition probabilities $\overline{p}(t, x, dy)$ satisfying the backward and forward equations for these parameters (see (2.63) in Chapter 2). Let A denote the infinitesimal generator of the Markov process $\{X_t\}$, and let $p(t; x, dy)$ be its transition probability.

For $t > 0$, $C \in \mathcal{S}$, the backward and forward equations take the form

$$
\begin{aligned}
[B]: \frac{\partial p(t;x,C)}{\partial t} &= -\lambda(x)p(t;x,C) + \int_S p(t;x_1,C)\lambda(x)k(x,dx_1),\\
[F]: \frac{\partial p(t;x,C)}{\partial t} &= -\int_C \lambda(x_1)p(t;x,dx_1) + \int_S \lambda(x_1)p(t;x,dx_1)k(x_1,C),
\end{aligned}
\tag{4.6}
$$

with

$$
p(0^+;x,dx_1) = \delta_{\{x\}}(dx_1),\ x \in S. \tag{4.7}
$$

These may be expressed in equivalent *mild* or *integrated* form as follows (Exercise 4): For $C \in \mathcal{S}$, $t > 0$,

$$
\begin{aligned}
[B]: p(t;x,C) &= e^{-\lambda(x)t}\delta_{\{x\}}(C) + \lambda(x)\int_0^t \int_S e^{-\lambda(x)t_1}k(x,dx_1)p(t-t_1;x_1,C)dt_1,\\
[F]: p(t;x,C) &= e^{-\lambda(x)t}\delta_{\{x\}}(C) + \int_0^t \int_S \lambda(x_1)\widetilde{k}(t_1;x_1,C)p(t-t_1;x,dx_1)dt_1,
\end{aligned}
\tag{4.8}
$$

where

$$
\widetilde{k}(t_1;x_1,C) = \int_C e^{-\lambda(x_2)t_1}k(x_1,dx_2),\ x_1 \in S, C \in \mathcal{S}. \tag{4.9}
$$

Remark 4.2 Let $f = \mathbf{1}_C$. Then, (4.6[B]) is the backward equation $\frac{d}{dt}T_t f(x) = (AT_t f)(x)$, formally applied to f. Similarly, (4.6[F]) is the forward equation $\frac{d}{dt}T_t f(x) = (T_t Af)(x)$ applied to $f = \mathbf{1}_C$. However, the forward equation is meant generally to describe the evolution of an initial probability distribution $\mu(dx) = f(x)\nu(dx)$, say. Applied to such a function f, the forward equation looks at the evolution of its density. More generally, if $\mu(dx)$ is a probability measure on $(S, \mathcal{S})$, then defining

$$
\mu(t,C) = \int_S \mu(dx)p(t;x,C), \quad t \ge 0, C \in \mathcal{S};
$$

then integrating both sides of (4.6[F]) with respect to $\mu(dx)$, the forward equations take a form familiar in physics and referred to as the *Fokker-Planck*, or *continuity* equation, governing how the rate at which probability mass in a region C of the state space changes in time as the result of spatial movement of a jump Markov process. Namely, for $C \in \mathcal{S}$, $t > 0$,

$$
\begin{aligned}
\frac{\partial \mu(t,C)}{\partial t} &= -\int_C \lambda(x_1)\mu(t,dx_1) + \int_S \lambda(x_2)k(x_2,C)\mu(t,dx_2)\\
&= -\int_C \lambda(x)k(x,C^c)\mu(t,dx) + \int_{C^c} \lambda(x)k(x,C)\mu(t,dx).
\end{aligned}
\tag{4.10}
$$

The steady state $\mu(t, C) = \mu(C), t \geq 0$, if there is one, corresponds to a choice of $\mu(dx)$ such that

$$\int_C \lambda(x)k(x, C^c)\mu(dx) = \int_{C^c} \lambda(x)k(x, C)\mu(dx), \quad C \in \mathcal{S}.$$

We will proceed in two steps: (i) Show that there is a smallest nonnegative solution to [B] and [F], in a sense to be defined, and (ii) show that this solution satisfies the Chapman–Kolmogorov equations, i.e., semigroup property

$$\begin{aligned} p(s+t; x, C) &= \int_S p(s; x, dx_1)p(t; x_1, C) \\ &= \int_S p(t; x, dx_1)p(s; x_1, C), \ x \in S, s, t \geq 0, \end{aligned} \tag{4.11}$$

that defines a corresponding Markov process. The approach to (i) was introduced by Feller (1940) based on a *method of successive approximations* that may be interpreted as *truncating*[1] the number of jumps to occur prior to time t.

Proposition 4.3 (Feller's Method of Successive Approximations) *Given any infinitesimal parameters $\lambda(x)$, $k(x, dy)$, there exists a smallest nonnegative solution $\overline{p}(t; x, dy)$ of the backward equations [B] satisfying the initial condition (4.7). Moreover, $\overline{p}(t; x, dy)$ solves the forward equations [F] with (4.7), as well.*

Proof To solve [B], start with the *first approximation*

$$p^{(0)}(t; x, dx_1) = e^{-\lambda(x)t}k^{(0)}(x, dx_1) \qquad (x \in S, t \geq 0) \tag{4.12}$$

where

$$k^{(0)}(x, dx_1) = \delta_{\{x\}}(dx_1), \tag{4.13}$$

and compute *successive approximations*, recursively for $n = 1, 2, \ldots, C \in \mathcal{S}$, by

$$p^{(n)}(t; x, C) = p^{(0)}(t; x, C) + \int_0^t \int_S \lambda(x)e^{-\lambda(x)(t-t_1)}p^{(n-1)}(t_1; x_1, C)k(x, dx_1)dt_1. \tag{4.14}$$

Then, $p^{(0)}(t; x, C) \leq p^{(1)}(t; x, C)$ and, by induction,

$$p^{(n-1)}(t; x, C) \leq p^{(n)}(t; x, C), \quad n = 1, 2, \ldots.$$

[1] This method is also effective in the analysis of mild forms of the incompressible Navier-Stokes equations in Fourier space by (semi-Markov) probabilistic methods introduced by Le Jan and Sznitman (1997); also see Bhattacharya and Waymire (2021) and Dascaliuc et al. (2023a,b).

Thus, for any $x \in S, C \in \mathcal{S}$,

$$\overline{p}(t; x, C) = \lim_{n \to \infty} p^{(n)}(t; x, C) \tag{4.15}$$

exists and

$$0 \le \overline{p}(t; x, C) \le \overline{p}(t; x, S) \le 1, \quad x \in S, C \in \mathcal{S}, t \ge 0.$$

This latter upper bound follows from $p^{(0)}(t; x, S) \le e^{-\lambda(x)t}$, so that $p^{(1)}(t; x, S) \le 1$. Thus, inductively,

$$p^{(n+1)}(t; x, S) \le e^{-\lambda(x)t} + \int_0^t \lambda(x) e^{-\lambda(x)s} ds = 1.$$

That $\overline{p}(t; x, C)$ solves the backward equations follows from Lebesgue's monotone convergence theorem. To prove minimality, suppose that $g(t; x, C), x \in S, C \in \mathcal{S}$, is another solution to [B]. Then, $g(t; x, C) \ge p^{(0)}(t; x, C)$, and inductively comparing [B] to (4.14), $g(t; x, C) \ge p^{(n)}(t; x, C)$ for all $n \ge 1$. Minimality follows by passing to the limit as $n \to \infty$. To prove that the forward equations [F] are also satisfied for $\overline{p}(t; x, dx_1)$, we apply a similar truncation of the number of jumps prior to time t but consider the recursion derived in terms of the time of the last jump (see Remark 4.3). Using the same notation for the new recursive probabilities, start the recursion with (4.7), but modify the recursion for $C \in \mathcal{S}$, $n \ge 1$, by

$$p^{(n)}(t; x, C) = p^{(0)}(t; x, C) + \int_0^t \int_S \lambda(x_1) \widetilde{k}(t - t_1; x_1, C) p^{(n-1)}(t_1; x, dx_1) dt_1. \tag{4.16}$$

Again, while the same notations for the recursions (4.14) and (4.16) are used, they define distinct recursions. As above, one may check, inductively, that $0 \le p^{(n)}(t; x, C) \le p^{(n+1)}(t; x, C)$ for all $x \in S, C \in \mathcal{S}, t \ge 0, n = 0, 1, \ldots$. Thus,

$$\phi(t; x, C) = \lim_{n \to \infty} p^{(n)}(t; x, C)$$

exists and solves the forward equation [F] by the monotone convergence theorem. To prove that, in fact,

$$\phi(t; x, dy) = \overline{p}(t; x, dy),$$

we show by induction that, in fact,

$$\phi^{(n)}(t; x, C) = \beta^{(n)}(t; x, C), \quad x \in S, t \ge 0, C \in \mathcal{S},\ n = 0, 1, 2 \ldots, \tag{4.17}$$

where $\beta^{(n)}(t)$ and $\phi^{(n)}(t)$ denote the nth approximate solutions to the backward and forward recursions, respectively. The case $n = 0$ is true by definition. Assume that (4.17), [B] and [F], hold for $n \leq m$. Then, starting with [F] and using the induction hypothesis, one has

$$
\begin{aligned}
&\phi^{(m+1)}(t; x, C) \\
&\quad = \phi^{(0)}(t; x, C) + \int_0^t \int_S \int_C \lambda(x_1) e^{-\lambda(x_2)(t-t_1)} k(x_1, dx_2) \beta^{(m)}(t_1; x, dx_1)) dt_1 \\
&\quad = \phi^{(0)}(t; x, C) + \int_0^t \int_S \int_C e^{-\lambda(x_2)(t-t_1)} k(x_1, dx_2) \delta_{\{x\}}(dx_1) \lambda(x) e^{-\lambda(x)t_1} dt_1 \\
&\qquad + \int_0^t \int_S \int_C \int_0^{t-t_1} \int_S e^{-\lambda(x_2)t_1} \lambda(x_1) k(x_1, dx_2) e^{-\lambda(x)t_2} \lambda(x) k(x, dx_3) \\
&\qquad\qquad \times \beta^{(m-1)}(t - t_1 - t_2; x_3, dx_1) dt_2 dt_1 \\
&\quad = \phi^{(0)}(t; x, C) + \int_0^t \int_C e^{-\lambda(x_2)(t-t_1)} \lambda(x) e^{-\lambda(x)t_1} k(x, dx_2) dt_1 \\
&\qquad + \int_0^t \int_S \int_C \int_S \int_{t_1}^t e^{-\lambda(x_2)t_1} e^{-\lambda(x)(t_2-t_1)} \lambda(x_1) k(x_1, dx_2) \lambda(x) k(x, dx_3) \\
&\qquad\qquad \times \beta^{(m-1)}(t - t_2; x_3, dx_1) dt_2 dt_1 \\
&\quad = F_1 + F_2 + F_3.
\end{aligned} \tag{4.18}
$$

On the other hand, starting with [B], the induction hypothesis yields

$$
\begin{aligned}
&\beta^{(m+1)}(t; x, C) \\
&\quad = \beta^{(0)}(t; x, C) + \int_0^t \int_S e^{-\lambda(x)t_1} \lambda(x) k(x, dx_1) \phi^{(m)}(t - t_1; x_1, C) dt_1 \\
&\quad = \beta^{(0)}(t; x, C) + \int_0^t \int_S e^{-\lambda(x)t_1} \lambda(x) k(x, dx_1) e^{-\lambda(x_1)(t-t_1)} \delta_{x_1}(C) dt_1 \\
&\qquad + \int_0^t \int_S \int_0^t \int_S \int_C e^{-\lambda(x)t_1} \lambda(x) k(x, dx_1) \lambda(x_2) e^{-\lambda(x_3)t_2} k(x_2, dx_3) \\
&\qquad\qquad \times \phi^{(m-1)}(t - t_1 - t_2; x_1, dx_2) dt_2 dt_1 \\
&\quad = \beta^{(0)}(t; x, C) + \int_0^t \int_C e^{-\lambda(x)t_1} \lambda(x) k(x, dx_1) e^{-\lambda(x_1)(t-t_1)} dt_1 \\
&\qquad + \int_0^t \int_S \int_C \int_S \int_{t_1}^t e^{-\lambda(x)t_1} \lambda(x) k(x, dx_1) \lambda(x_2) e^{-\lambda(x_3)(t_2-t_1)} k(x_2, dx_3) \\
&\qquad\qquad \times \phi^{(m-1)}(t - t_2, x_1, dx_2) dt_2 dt_1 \\
&\quad = B_1 + B_2 + B_3.
\end{aligned} \tag{4.19}
$$

One readily sees by definition that $F_1 = B_1$. Also $F_2 = B_2$ involves renaming integration variable x_2 as x_1 and a change of variables $t - t_1 \to t_1$. Checking that $F_3 = B_3$ is similarly accomplished, using the induction hypothesis, renaming space variables, and change of variables in time, respectively. Thus, one obtains term by term agreement between the right sides of (4.18) and (4.19). ■

Remark 4.3 Note that $p^{(0)}(t; x, dx_1)$ is the distribution of the process starting at x, until the time of the first jump. Similarly, $p^{(1)}(t; x, dx_1)$ is the distribution of the process until the time of the second jump, and so on.

Theorem 4.4 *Given any infinitesimal parameters $\lambda(x), k(x, dy)$, there exists a smallest nonnegative solution $\overline{p}(t; x, dy)$ of the backward equations [B] and the forward equations [F], for the initial condition (4.7). This solution satisfies the Chapman–Kolmogorov semigroup equations (4.11) and the Feller property $\lim_{t\to 0^+} \overline{p}(t; x, dx_1) = \delta_{\{x\}}(dx_1)$, $x \in S$. However,*

$$\int_S \overline{p}(t; x, dx_1) \leq 1 \quad \textit{for all } x \in S, \ \textit{all} \ \ t \geq 0. \tag{4.20}$$

In case equality holds in this last inequality (4.20) for all $x \in S$ and $t \geq 0$, there does not exist any other nonnegative solution of [B], nor [F], and satisfying the initial condition (4.7).

Proof The existence of the minimal solution has been settled. Our task is to prove the Chapman–Kolmogorov equations for the the minimal solution $\overline{p}(t; x, dy)$ to the backward equations for these parameters. This is accomplished as follows. Define, for $C \in \mathcal{S}$,

$$d^{(n)}(t; x, C) = \begin{cases} \delta_x(C)e^{-\lambda(x)t} & \text{if } n = 0 \\ \beta^{(n)}(t; x, C) - \beta^{(n-1)}(t; x, C). & \text{if } n \geq 1. \end{cases} \tag{4.21}$$

The strategy is to prove for $s, t \geq 0$,

$$\beta^{(n)}(t + s; x, C) = \int_S \sum_{m=0}^{n} d^{(m)}(s; x, dx_1)\beta^{(n-m)}(t : x_1, C). \tag{4.22}$$

Then, the Chapman–Kolmogorov equations will follow in the limit as $n \to \infty$ using monotone convergence and the telescoping sum

$$\overline{p}(t; x, C) = \sum_{m=0}^{\infty} d(t; x, C), \quad x \in S, t \geq 0.$$

The proof of (4.22) is by induction. It is obvious for $n = 0$. To simplify notation for the backward iteration, let

$$\varepsilon(t; x, C) = \begin{cases} \lambda(x)e^{-\lambda(x)}k(x, C) & \text{if } \lambda(x) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Then, (4.14) can be expressed as

$$\beta^{(n+1)}(t; x, C) = e^{-\lambda(x)t}\delta_{\{x\}}(C) + \int_0^t \int_S \varepsilon(t - t_1, x, dx_1)\beta^{(n)}(t_1, x_1, C), n \geq 0.$$

As a lemma, one may also check by induction and the recursion for [B] (Exercise 5),

$$d^{(m+1)}(t; x, C) = \int_0^t \int_S \varepsilon(t_1; x, dx_1)d^{(m)}(t - t_1; x_1, C)dt_1 \; m \geq 0. \tag{4.23}$$

$$\begin{aligned}
&\beta^{(n+1)}(s + t; x, C) \\
&\quad = \delta_x(C)e^{-\lambda(x)(s+t)} + \int_0^{s+t} \int_S \varepsilon(t_2; x, dx_2)\beta^{(n)}(s + t - t_2; x_2, C)dt_2 \\
&\quad = \delta_x(C)e^{-\lambda(x)(t+s)} + \int_s^{t+s} \int_S \varepsilon(t_2; x, dx_2)\beta^{(n)}(s + t - t_2; x_2, C)dt_1 \\
&\qquad + \int_0^s \int_S \varepsilon(t_2; x, dx_2)\beta^{(n)}(s + t - t_2; x_2, C)dt_2 \\
&\quad = \delta_x(C)e^{-\lambda(x)(t+s)} + \int_s^{t+s} \int_S \varepsilon(t_2; x, dx_2)\beta^{(n)}(s + t - t_2; x_2, C)dt_2 \\
&\qquad + \int_0^s \int_S \varepsilon(t_2; x, dx_2)\beta^{(n)}(s + t - t_2; x_2, C)dt_2 \\
&\quad = \delta_x(C)e^{-\lambda(x)(t+s)} + \int_S \delta_x(dx_1)e^{-\lambda(x)s} \int_0^t \int_S \varepsilon(t - t_1; x_1, dx_2)\beta^{(n)}(t_1; x_2, C)dt_1 \\
&\qquad + \int_0^s \int_S \varepsilon(t_2; x, x_2)\beta^{(n)}(s + t - t_2; x_2, C)dt_2 \\
&\quad = \int_S \delta_x(dx_1)e^{-\lambda(x)s}\{\delta_{x_1}(C)e^{-\lambda(x_1)t} + \int_0^t \int_S \varepsilon(t - t_1; x_1, dx_2)\beta^{(n)}(t_1; x_2, C)dt_1\} \\
&\qquad + \int_S \sum_{m=0}^n \{\int_0^s \int_S \varepsilon(t_2; x, dx_2)d^{(m)}(s - t_2; x_1, dx_2)\}dt_2\beta^{(n-m)}(t; x_1, C) \\
&\quad = \int_S \delta_x(dx_1)e^{-\lambda(x_1)s}\beta^{(n+1)}(t; x_1, C)) + \int_S \sum_{m=0}^n d^{(m+1)}(s; x, dx_1)\beta^{(n-m)}(t; x_1, C)
\end{aligned}$$

$$= \int_S d^{(0)}(s; x, dx_1)\beta^{(n+1)}(t; x_1, C) + \int_S \sum_{m=1}^{n+1} d^{(m)}(s; x, dx_1)\beta^{(n+1-m)}(t; x_1, C)$$

$$= \int_S \sum_{m=0}^{n+1} d^{(m)}(s; x, dx_1)\beta^{(n+1-m)}(t; x_1, C) \tag{4.24}$$

This completes the induction, and the Chapman–Kolmogorov equations now follow in the limit as $n \to \infty$. ■

To allow for the possibility that $\overline{p}(t; x, S) < 1$, one may adjoin a fictitious state, denoted s_∞, to S and extend $\overline{p}(t; x, S) < 1$ to

$$\overline{S} = S \cup \{s_\infty\}, \quad \overline{\mathcal{S}} = \mathcal{S} \vee \sigma(\{s_\infty\}),$$

by

$$\overline{p}(t; x, \{s_\infty\}) = 1 - \overline{p}(t; x, S),\ x \in S, \quad \overline{p}(t; s_\infty, \{s_\infty\}) = 1.$$

Then, using the Kolmogorov extension theorem, one may construct a Markov process $\overline{X} = \{\overline{X}_t : t \geq 0\}$ on a probability space $(\Omega, \mathcal{F}, P)$, having state space $(\overline{S}, \overline{\mathcal{S}})$ and transition probabilities $\overline{p}(t; x, dy)$. Alternatively, one may use Kolmogorov's extension theorem to construct a sequence of nonnegative mean one exponentially distributed random variables $\{\overline{T}_n : n \geq 0\}$ and, independently, a Markov chain $\{Y_n : n = 0, 1, \dots\}$ on S with transition probability kernel $k(x, dy)$. Let

$$T_m = \lambda^{-1}(Y_m)\overline{T}_m,\ m \geq 1.$$

Then, $\overline{X}$ may be constructed as

$$\overline{X}_t = \begin{cases} Y_{N_t} & \text{for finite } N_t \\ s_\infty & \text{otherwise} \end{cases} \tag{4.25}$$

where

$$N_t = \inf\{n \geq 0 : \sum_{m=0}^{n} T_m > t\}. \tag{4.26}$$

That the two processes are stochastically equivalent follows by checking that their finite dimensional distributions coincide (Exercise 3).

Definition 4.1 The Markov process, $\overline{X}$ or its restriction to $0 \leq t < \zeta$, where

$$\zeta := \inf\{t \geq 0 : \overline{X}_t \notin S\},$$

is referred to as the *minimal process*. The random variable ζ is referred to as the *explosion time*.

The connection between $p^{(n)}(t; x, dy)$ and $\overline{p}(t; x, dy)$ and the minimal process $\overline{X}$ are as follows.

Proposition 4.5 *For $x \in S$, $C \in \mathcal{S}$, $t \geq 0$,*

a. $p^{(n)}(t; x, C) = P_x(\overline{X}_t \in C, N_t < n), \quad n = 1, 2, \ldots.$
b. $\overline{p}(t; x, C) = P_x(\overline{X}_t \in C, \zeta > t).$
c. $P_x(\zeta > t) = \overline{p}(t; x, S).$

Proof One has, conditioning on T_0 and Y_1

$$\begin{aligned}
&P_x(\overline{X}_t \in C, N_t < n+1) \\
&\quad = \mathbb{E}_x P_x(\overline{X}_t \in C, N_t < n+1|\sigma(Y_1, T_0)) \\
&\quad = \mathbb{E}_x\{\mathbf{1}_C(x)\mathbf{1}_{[T_0>t]}\} + \mathbb{E}_x\{P_x(\overline{X}_{t-T_0} \in C, N_{t-T_0} < n|\sigma(Y_1, T_0))\mathbf{1}[T_0 \leq t]\} \\
&\quad = \delta_{\{x\}}(C)e^{-\lambda(x)t} + \int_0^t \lambda(x)e^{-\lambda(x)t_1}\int_S k(x, dx_1)P_x(\overline{X}_{t-t_1} \in dx_1, N_{t-t_1} < n)dt_1.
\end{aligned} \tag{4.27}$$

Thus, $p^{(n)}(t; x, C)$ and $P_x(\overline{X}_t \in C, N_t < n)$ are solutions to the same recursion (4.14) with common initialization $\delta_{\{x\}}(C)e^{-\lambda(x)t}$ for $n = 0$ and, by induction, must coincide; (b) follows in the limit as $n \to \infty$; since $[\zeta > t] = [N_t < \infty]$; and (c) follows since the added state s_∞ to S is absorbing for $\{\overline{X}_t : t \geq 0\}$. In particular, $[\overline{X}_t \in S] = [\zeta > t]$ ■

Remark 4.4 Theorem 4.1 and Corollary 4.2 describe the structure of a jump Markov process with infinitesimal parameters $\lambda(x)$, $p(x, dy)$ $(x \in S)$. If conversely, one can prove that the process Y, say, in Corollary 4.2 is Markov, then the rather ungainly proof of the Chapman–Kolmogorov property would be unnecessary; for the process Y satisfies the properties in Proposition 4.5 of the minimal process. The proof of the Markov property of Y is hinted at for a countable state space S in Bhattacharya and Waymire (2009), Theorem 5.4, pp. 278–279.

Proposition 4.6 *The minimal process is a conservative pure jump process if and only if*

$$P_x(\sum_{n=0}^{\infty}\frac{1}{\lambda(Y_n)} < \infty) = 0 \quad \textit{for all } x \in S.$$

Proof Assume first that $\sum_{k=0}^{\infty}\frac{1}{\lambda(Y_k)} = \infty$ a.s. Recall that

$$\zeta = \lim_{n\to\infty} \tau_n = \lim_{n\to\infty} \sum_{k=0}^{n} T_k$$

where conditionally given $\sigma(Y_0, Y_1, \ldots)$ and the random variables $T_0, T_1, \ldots$ are independent and exponentially distributed with respective parameters $\lambda(Y_0), \lambda(Y_1), \ldots$. By the dominated convergence theorem, one has for each $x \in S$,

$$\begin{aligned}
\mathbb{E}_x e^{-\zeta} &= \lim_{n\to\infty} \mathbb{E}_x e^{-\Sigma_{k=0}^n T_k} = \lim_{n\to\infty} \mathbb{E}_x \mathbb{E}\{e^{-\Sigma_{k=0}^n T_k} \mid \sigma(Y_0, Y_1, \ldots)\} \\
&= \lim_{n\to\infty} \mathbb{E}_x \prod_{k=0}^{n} \frac{\lambda(Y_k)}{1+\lambda(Y_k)} \\
&= \mathbb{E}_x \lim_{n\to\infty} \prod_{k=0}^{n} \frac{\lambda(Y_k)}{1+\lambda(Y_k)} = \mathbb{E}_x \lim_{n\to\infty} \frac{1}{\prod_{k=0}^n (1+\frac{1}{\lambda(Y_k)})}.
\end{aligned}$$

Now,

$$\begin{aligned}
\log\Big(\frac{1}{\prod_{k=0}^n (1+\frac{1}{\lambda(Y_k)})}\Big) &= -\sum_{k=0}^{n} \log\Big(1+\frac{1}{\lambda(Y_k)}\Big) \\
&\le -\sum_{k=0}^{n} \frac{\frac{1}{\lambda(Y_k)}}{1+\frac{1}{\lambda(Y_k)}} = -\sum_{k=0}^{n} \frac{1}{1+\lambda(Y_k)}
\end{aligned}$$

since $\log(1+x) \ge x/(1+x)$ for $x \ge 0$. In particular, $\log(\frac{1}{\prod_{k=0}^n (1+\frac{1}{\lambda(Y_k)})}) \to -\infty$ a.s. as $n \to \infty$. Thus, $\prod_{k=0}^n (1+\frac{1}{\lambda(Y_k)}) \to \infty$ a.s. as $n \to \infty$. Again, by the dominated convergence theorem, one has $\mathbb{E}_x e^{-\zeta} = 0$. Thus, $\zeta = +\infty$ a.s. Conversely if $\zeta = \infty$ a.s., then these calculations show that $\mathbb{E} e^{-\zeta} = 0$ and, therefore almost surely,

$$\prod_{k=0}^{\infty} \Big(1+\frac{1}{\lambda(Y_k)}\Big) = \infty.$$

Thus, $\sum_{k=0}^{\infty} \frac{1}{\lambda(Y_k)} = \infty$ a.s. ■

Once again, we arrive at (see Remark 4.1),

Corollary 4.7 (Bounded Rates) *If* $\sup_{x\in S} \lambda(x) < \infty$*, then the minimal process is nonexplosive.*

Corollary 4.8 *If the embedded spatial process* $\{Y_k\}_{k=0}^{\infty}$ *is pointwise recurrent, then the minimal process is nonexplosive.*

The following theorem and proposition of Reuter (1957) also provide useful approaches to the determination of explosion/nonexplosion events.

Theorem 4.9 (Reuter's Conditions for Explosion) *Let $\lambda(x)$, $k(x, dy)$ be given infinitesimal rates for $\overline{X}$. Define a linear operator*

$$Af(x) = \lambda(x) \int_S \{f(y) - f(x)\} k(x, dy), \quad x \in S.$$

Then, $\overline{X}$, starting from x_0, is explosive if and only if there is a bounded, nonnegative function u and real number $\gamma > 0$, such that $u(x_0) > 0$ for some x_0, and

$$Au(x) \geq \gamma u(x), \quad \forall x \in S.$$

Proof Define a bounded, nonnegative function

$$u_\gamma(x) = \mathbb{E}_x e^{-\gamma\zeta}, x \in S.$$

Then, $P_{x_0}(\zeta < \infty) > 0$ if and only if $u_\gamma(x_0) > 0$. Moreover, conditioning on the process up to the time τ_1 of the first jump,

$$\begin{aligned} u_\gamma(x) &= \mathbb{E}_x \left(\mathbb{E}_x(e^{-\gamma\zeta} | \mathcal{F}_{\tau_1})\right) \\ &= \mathbb{E}_x \left(e^{-\gamma\tau_1} u_\gamma(X_{\tau_1})\right) \\ &= \frac{\lambda(x)}{\lambda(x) + \gamma} \int_S u_\gamma(y) k(x, dy). \end{aligned} \tag{4.28}$$

Equivalently,

$$Au_\gamma(x) = \gamma u_\gamma(x), \quad x \in S.$$

If $P_{x_0}(\zeta < \infty) > 0$, then $u_\gamma(x_0) > 0$. Thus, necessity of the condition for explosion with positive probability is proved. For sufficiency, suppose that u is a nonnegative bounded function, say $u \leq B$, such that for some $\gamma > 0$,

$$Au(x) \geq \gamma u(x), \quad \forall\, x \in S. \tag{4.29}$$

Define a martingale by

$$\begin{aligned} M_t &= e^{-\gamma t} u(\overline{X}_t) - \int_0^t e^{-\gamma s}(Au - \gamma u)(\overline{X}_s) ds \\ &= e^{-\gamma t} u(\overline{X}_t) - I_t, \quad t \geq 0, \end{aligned} \tag{4.30}$$

where $t \to I_t$ is a nondecreasing function by (4.29). In particular, therefore, $e^{-\gamma t} u(\overline{X}_t) = M_t + I_t, t \geq 0$, is a uniformly bounded submartingale. Let τ_n denote the time of the nth jump. By Doob's stopping time theorem, $e^{-\gamma\tau_n} u(\overline{X}_{\tau_n}), n \geq 1$, is a submartingale and

$$\mathbb{E}_{x_0} e^{-\gamma \cdot 0} u(\overline{X}_0) \leq \mathbb{E}_{x_0} e^{-\gamma \tau_n} u(\overline{X}_{\tau_n}) \leq B\mathbb{E}_{x_0} e^{-\gamma \tau_n} \leq BP_{x_0}(\zeta < \infty). \tag{4.31}$$

Letting $n \to \infty$, one has

$$u(x_0) \leq B\mathbb{E}_{x_0} e^{-\gamma \zeta}.$$

Thus, $P_{x_0}(\zeta < \infty) \geq B^{-1}u(x_0) > 0$. ■

Proposition 4.10 *Let $\lambda(x)$, $k(x, dy)$ be given infinitesimal rates for $\overline{X}$. Assume that there is a nonnegative function u with the property that for any sequence $\{x_n\}$ in S such that $\lim_n \lambda(x_n) = \infty$, one also has $\lim_n u(x_n) = \infty$, and for which there is a real number $\gamma > 0$ such that*

$$Au(x) \leq \gamma u(x), \quad \forall\, x \in S,$$

where

$$Af(x) = \lambda(x) \int_S \{f(y) - f(x)\}k(x, dy), \quad x \in S.$$

Then, $\overline{X}$ is not explosive.

Proof Since $e^{-\gamma t}u(X_t), t \geq 0$, is a nonnegative supermartingale, one has for any $T > 0$,

$$\mathbb{E}_x e^{-\gamma(\tau_n \wedge T)} u(\overline{X}_{\tau_n \wedge T}) \leq u(x), \quad \forall x \in S,$$

where, again, τ_n denotes the time of the nth jump. Letting $T \to \infty$, one has by Fatou's lemma that

$$\mathbb{E}_x e^{-\gamma \tau_n} u(\overline{X}_{\tau_n}) \leq u(x), \quad \forall x \in S. \tag{4.32}$$

Now, by the same argument as used in the proof of Proposition 4.6, $P(\zeta = \lim_n \tau_n < \infty) > 0$ implies that $P(\sum_{n=1}^{\infty} \frac{1}{\lambda(\overline{X}_{\tau_n})} < \infty) > 0$. Thus, $\lim_n \lambda(\overline{X}_{\tau_n}) \to \infty$ is an event with positive probability. In view of the condition on u in the statement of the theorem, the event that $u(\overline{X}_{\tau_n}) \to \infty$ has positive probability as well. But (4.32) makes this impossible, i.e., an event with probability zero. ■

Corollary 4.11 *Let $\lambda(x)$, $k(x, dy)$ be given infinitesimal rates. If $\gamma = \sup_x \int_S \lambda(y)k(x, dy) < \infty$, $x \in S$, then $\overline{X}$, starting at $x \in S$, is nonexplosive.*

Proof Define $u(x) = \lambda(x)$, $x \in S$. Then,

$$Au(x) = \lambda(x) \int_S \{\lambda(y) - \lambda(x)\}k(x, dy)$$

$$\leq \lambda(x) \int_S \lambda(y) k(x, dy)$$
$$\leq \gamma \lambda(x) = \gamma u(x).$$

■

In the case $P_x(\zeta = \infty) < 1$, there are various ways of continuing $\overline{X}$ for $t \geq \zeta$ so as to preserve the Markov property and the backward equations for the continued process $\{\widetilde{X}_t : t \geq 0\}$. An illustration is provided in Example 1. An important consequence is the role that these continuations play in the *nonuniqueness* of nonnegative solutions to the backward equations.

The case of a countable state space S plays a large role in the present context. Continuous parameter jump Markov chains are often defined in terms of an array

$$Q = ((q_{ij}))_{i,j \in S}$$

of parameters such that

$$q_{ij} \geq 0, i \neq j, \text{ and } \sum_j q_{ij} = 0, \text{ for all } i \in S. \tag{4.33}$$

Q then specifies the infinitesimal parameters in the case $q_{ii} \neq 0$, by

$$\lambda_i = -q_{ii}, \text{ and } k(i, \{j\}) = \frac{q_{ij}}{\lambda_i}, \ i \neq j, \ k(i, \{i\}) = 0.$$

If $\lambda_i = -q_{ii} = 0$, then the convention of treating i as an absorbing state is invoked, i.e., $k(i, \{i\}) = 1$. The *infinitesimal rates* are given by an array $Q = ((q_{ij}))$ satisfying (4.33). Q is also referred to as the Q-matrix in this context.

The role of the embedded Markov chain is particularly revealing in considerations of transience and recurrence properties.

Definition 4.2 Let $\{X_t : t \geq 0\}$ be a continuous parameter Markov chain on a countable state space S. A state $j \in S$ is recurrent if $P_j(\sup\{t \geq 0 : X_t = j\} = \infty) = 1$. Otherwise, $j \in S$ is said to be transient.

Theorem 4.12 *Let $\{X_t : t \geq 0\}$ be a continuous parameter Markov chain on a countable state space S. A state $j \in S$ is transient or recurrent for $\{X_t\}$ according to whether j is transient or recurrent for the embedded discrete parameter Markov chain $\{Y_n : n = 0, 1, \ldots\}$.*

Proof The result is trivially true in the case of absorption and left aside. If $j \in S$ is recurrent for $\{X_t\}$, then $\{Y_n\}$ will return to j at infinitely many jump times of $\{X_t\}$ with probability one, and j is thus recurrent for $\{Y_n\}$. Conversely, if j is recurrent fro $\{Y_n\}$, then under P_j, the successive return times

$$\tau_j^{(r)} = \inf\{t \geq \tau_j^{(r-1)} : X_t = j\}, \quad r = 1, 2, \ldots,$$

where $\tau_j^{(0)}$ is the initial holding time in state j, have i.i.d. positive increments $\tau_j^{(1)}, \tau_j^{(r)} - \tau_j^{(r-1)}, r \geq 2$. In particular, therefore, P_j-a.s.,

$$\tau_j^{(n)} = \tau_j^{(1)} + \sum_{r=2}^{n} (\tau_j^{(r)} - \tau_j^{(r-1)}) \to \infty.$$

Thus, j is recurrent for $\{X_t\}$, ∎

In view of the almost sure dichotomy between transient and recurrent states of discrete parameter Markov chains (Exercise 16), one has the following.

Corollary 4.13 *A state $j \in S$ is transient for $\{X_t\}$ if and only if $P_j(\sup\{t \geq 0 : X_t = j\} < \infty) = 1$. In this case, j is transient for the embedded Markov chain.*

Example 1 (Continuous Parameter Birth–Death Chains) For given sequences of nonnegative numbers $\{\beta_i\}_{i \in \mathbb{Z}}$, $\{\delta_i\}_{i \in \mathbb{Z}}$, the minimal Markov chain on $S = \mathbb{Z}$ specified by the infinitesimal rates

$$q_{i,i+1} = \beta_i,\ q_{i,i-1} = \delta_i,\ q_{ii} = -\lambda_i := -(\beta_i + \delta_i),\ q_{ij} = 0, |i - j| \geq 2, i, j \in \mathbb{Z}, \tag{4.34}$$

is referred to as a (continuous parameter) *birth–death chain.*

The further specialization $\delta_i = 0$ for all $i \in \mathbb{Z}$ is referred to as a *pure birth* process, and $\beta_i = 0$ for all $i \in \mathbb{Z}$ defines the *pure death* process. The state space may naturally be further restricted to $S = \mathbb{Z}^+$ in cases $\beta_i = \delta_i = 0$ for $i \leq -1$. It follows from Proposition 4.6 that a pure birth process is nonexplosive if and only if

$$\sum_{n=1}^{\infty} \frac{1}{\lambda_n} = \infty.$$

Now, consider an explosive pure birth process, e.g., $\lambda_i = \frac{1}{i^2}, i \in S = \mathbb{Z}^+$. Let $\{\overline{X}_t : 0 \leq t < \zeta\}$ denote the minimal process on S starting at $i = 0$. Consider the extension $X_t^{(0)}, 0 \leq t \leq \zeta^{(0)} := \zeta$ of $\{\overline{X}_t : 0 \leq t < \zeta\}$ by an instantaneous return to zero, i.e.,

$$X_t^{(0)} = \overline{X}_t, 0 \leq t < \zeta, \quad X_\zeta^{(0)} = 0.$$

Next define

$$R_n := \sum_{j=0}^{n} \zeta^{(j)}, n \geq 0,\ R_{-1} = 0,$$

and let $\{X_t^{(n)} : 0 \le t < \zeta^{(n)}\}$, $n \ge 1$, be i.i.d. copies of $X_t^{(0)}$. Define

$$X_{R_{n-1}} = X_{R_n} = 0,$$

and

$$\widetilde{X}_t = \begin{cases} X_t^{(0)} & \text{if } 0 \le t \le \zeta^{(0)} \\ X_t^{(n)} & \text{if } R_{n-1} \le t \le R_n, n \ge 1. \end{cases} \tag{4.35}$$

Since $\zeta^{(n)}, n \ge 0$, are i.i.d. positive random variables, one has $R_n \to \infty$ a.s.. So $\widetilde{X}_t$ is defined for all $t \ge 0$. Clearly

$$P_0(\widetilde{X}_t = j) > P_0(\overline{X}_t = j) = \overline{p}(t; 0, \{j\}),$$

since $[\overline{X}_t = j] \subset [\widetilde{X}_t = j]$ is a proper subset. We leave it as Exercise 6 to check that the backward equations hold, but the forward equations fail. Note that the extension could have been made by an instantaneous return to any state $k \in \mathbb{Z}^+$, or for that may to a randomly selected state, independent of the process. Such fascinating considerations[2] lead to infinitely many solutions to the backward equations.

More generally, explosion starting in state j implies that j is transient. In view of Theorem 4.12, the criteria for transience or recurrence of a continuous parameter birth–death chain on $S = \{0, 1, \dots\}$ with reflecting boundary at 0 are the same as those for the discrete parameter embedded Markov chain, namely, convergence or divergence of the series $\sum_{n=1}^{\infty} \frac{\delta_1 \cdots \delta_j}{\beta_1 \cdots \beta_j}$ (Exercise 11 or see Bhattacharya and Waymire 2022). To derive necessary and sufficient conditions for explosion of the birth–death process on $S = \{0, 1, \dots\}$ with reflection at 0, one may apply Theorem 4.9 as follows.

Proposition 4.14 (Reuter) *Let $\{X_t : t \ge 0\}$ be a continuous parameter birth–death process on $S = \{0, 1, \dots\}$ with reflection at 0 and starting in state $j \in S$. Assume $\beta_j \delta_j > 0$, $j = 1, 2 \dots$, $\beta_0 = 1$, $\delta_0 = 0$, and $0 < \lambda_j < \infty$, $j \ge 0$. Then, $\{X_t\}$ is P_j-a.s. explosive for each $j \in S$ if and only if*

$$\sum_{n=1}^{\infty} \frac{\delta_1 \cdots \delta_n}{\beta_1 \cdots \beta_n} \sum_{j=1}^{n} \frac{\beta_1 \cdots \beta_j}{\delta_1 \cdots \delta_j} \frac{1}{\lambda_j \beta_j} < \infty, \quad \sum_{j=1}^{\infty} \frac{\delta_1 \cdots \delta_j}{\beta_1 \cdots \beta_j} < \infty.$$

Proof Suppose that u is a positive function, such that $Au(j) = \gamma u(j)$, $j \ge 0$, for $\gamma > 0$, where

$$Au(j) = \lambda_j \{\beta_j u(j+1) + \delta_j u(j-1) - u(j)\}, \quad j \ge 0 \ (u(-1) = 0).$$

---

[2] See Freedman (1983) for a more extensive treatment of such extensions.

Then, writing $\tilde{\beta}_j = \lambda_j \beta_j, \tilde{\delta}_j = \lambda_j \delta_j$, $j \geq 0$, $\lambda_j = \tilde{\beta}_j + \tilde{\delta}_j$, $\frac{\tilde{\delta}_j}{\tilde{\beta}_j} = \frac{\delta_j}{\beta_j}$,

$$\begin{aligned}
&\tilde{\beta}_0 u(1) = (\gamma + \tilde{\beta}_0) u(0) \\
&\tilde{\beta}_j u(j+1) + \tilde{\delta}_j u(j-1) = (\gamma + \tilde{\beta}_j + \tilde{\delta}_j) u(j), \ \ j \geq 1.
\end{aligned} \tag{4.36}$$

Then,

$$u(j+1) = \frac{(\gamma + \lambda_j) u(j) - \tilde{\delta}_j u(j-1)}{\tilde{\beta}_j}, \quad j \geq 0. \tag{4.37}$$

Letting $v(j+1) = u(j+1) - u(j)$, $j \geq 0$, and taking $u(0) = 1$ to standardize the solution, it follows that $v(1) = u(1) - 1 = \frac{\gamma}{\tilde{\beta}_0}$ and

$$v(j+1) = \frac{\delta_j}{\beta_j} v(j) + \frac{\gamma}{\tilde{\beta}_j} u(j), j \geq 0. \tag{4.38}$$

In particular, inductively one has $v(j) > 0$, $j \geq 0$, and therefore, u must be strictly increasing. Now, one may recursively compute

$$\begin{aligned}
v(j+1) &= \frac{\gamma}{\tilde{\beta}_j} u(j) + \frac{\delta_j}{\beta_j} \frac{\gamma}{\tilde{\beta}_{j-1}} u(j-1) + \cdots + \frac{\gamma}{\tilde{\beta}_0} \frac{\delta_j \cdots \delta_1}{\beta_j \cdots \beta_1} u(0) \\
&\geq \frac{\gamma}{\tilde{\beta}_j} + \frac{\delta_j}{\beta_j} \frac{\gamma}{\tilde{\beta}_{j-1}} + \cdots + \frac{\gamma}{\tilde{\beta}_0} \frac{\delta_j \cdots \delta_1}{\beta_j \cdots \beta_1}.
\end{aligned} \tag{4.39}$$

Now, $\lim_{j \to \infty} u(j) = \sum_{j=1}^{\infty} v(j) + 1 = \infty$ if $\sum_{n=1}^{\infty} \frac{\delta_1 \cdots \delta_n}{\beta_1 \cdots \beta_n} \sum_{j=1}^{n} \frac{\beta_1 \cdots \beta_j}{\delta_1 \cdots \delta_j} \frac{1}{\lambda_j \beta_j} = \infty$, i.e., u is an unbounded solution, and therefore ,$\{X_t\}$ is nonexplosive. On the other hand, again by monotonicity,

$$v(j) \leq \gamma \left(\frac{1}{\tilde{\beta}_j} + \frac{\delta_j}{\beta_j} \frac{1}{\tilde{\beta}_{j-1}} + \cdots + \frac{\delta_j \cdots \delta_1}{\beta_j \cdots \beta_1} \frac{1}{\tilde{\beta}_0}\right) u(j), \quad j \geq 1. \tag{4.40}$$

So,

$$\begin{aligned}
&u(j+1) \\
&\quad \leq \left(1 + \frac{\gamma}{\tilde{\beta}_j} + \frac{\delta_j}{\beta_j} \frac{\gamma}{\tilde{\beta}_{j-1}} + \cdots + \frac{\delta_j \cdots \delta_1}{\beta_j \cdots \beta_1} \frac{\gamma}{\tilde{\beta}_0}\right) u(j) \\
&\quad \leq \exp\left(\frac{\gamma}{\tilde{\beta}_j} + \frac{\delta_j}{\beta_j} \frac{\gamma}{\tilde{\beta}_{j-1}} + \cdots + \frac{\delta_j \cdots \delta_1}{\beta_j \cdots \beta_1} \frac{\gamma}{\tilde{\beta}_0}\right) u(j)
\end{aligned}$$

$$\vdots$$

$$\leq \exp\left(\sum_{k=0}^{j}\left\{\frac{1}{\tilde{\beta}_k}+\frac{\delta_k}{\beta_k}\frac{1}{\tilde{\beta}_{k-1}}+\cdots+\frac{\delta_k\cdots\delta_2}{\beta_k\cdots\beta_2}\frac{\gamma}{\tilde{\beta}_1}u(1)+\frac{\delta_k\cdots\delta_1}{\beta_k\cdots\beta_1}\frac{1}{\tilde{\beta}_0}\right\}\right)u(0).$$

Thus, if $\sum_{n=1}^{\infty}\frac{\delta_1\cdots\delta_n}{\beta_1\cdots\beta_n}\sum_{j=1}^{n}\frac{\beta_1\cdots\beta_j}{\delta_1\cdots\delta_j}\frac{1}{\lambda_j\beta_j} < \infty$ and $\sum_{k=1}^{\infty}\frac{\delta_k\cdots\delta_1}{\beta_k\cdots\beta_1} < \infty$, then u must be bounded, and $\{X_t\}$ explodes. ■

For another interesting application of Theorem 4.12, see Exercise 13.

Remark 4.5 Even in the case of countable state spaces, general continuous parameter Markov chains are capable of more exotic behavior than presented here, e.g., see Chung (1967) and Freedman (1983). The examples by Lévy (1951), Dobrushin (1956), Feller and McKean (1956), and Blackwell (1958) are especially instructive in identifying nuances of the general theory.

Exercises

1. *(Cauchy's Functional Equation)* Suppose that a positive right-continuous function φ on $[0,\infty)$ satisfies $\varphi(s+t)=\varphi(s)\varphi(t)$, for all $s,t\geq 0$. Prove that φ is of the form $\varphi(t)=e^{-\lambda t}$, for some $\lambda\in\mathbb{R}$. [*Hint*: Let $f=\log\varphi$. Then, f satisfies $f(s+t)=f(s)+f(t)$, $s,t\geq 0$, implying $f(mx)=mf(x)$, $f(\frac{m}{n}x)=\frac{m}{n}f(x)$ for all $m,n\in\mathbb{Z}_+$, $x\geq 0$ (noting that $f(x)=nf(\frac{1}{n}x)$). Now use right-continuity to prove $f(cx)=cf(x)$ for all $c>0$ and hence $f(x)=xf(1)=\theta x$ for all $x\geq 0$ and for some constant θ.]
2. Supply the details for the proof of Corollary 4.2. [*Hint*: Use the strong Markov property.]
3. Show that the two definitions of the minimal process $\overline{X}$ obtained from Kolmogorov extension theorem applied to $\overline{p}(t:x,dy)$ and (4.25), respectively, are stochastically equivalent.
4. Establish the equivalence of the integro-differential and mild forms of the backward and forward equations given in (4.8).
5. Give a proof of (4.23). [*Hint*: Use mathematical induction.]
6. Show that the transition probabilities of the instantaneous return to zero extension $\widetilde{X}$ of the minimal process fail to satisfy the forward equation, but not the backward equations. [*Hint*: For $n\geq 0$, show that $\tilde{p}_{00}(t)=\sum_{n=0}^{\infty}\tilde{P}_0(\widetilde{X}_t=0, N_t=n)=\sum_{n=0}^{\infty}\int_0^t\overline{p}_{0,0}(t-s)f_0^{*n}(s)ds$, where f_0 is the pdf of ζ, since the times between successive explosions are i.i.d. with pdf f_0 and $\overline{p}(t)=e^{-\lambda_0 t}$. Differentiate to see the forward equations fail. To show the backward equations in this case, write $f_0=e_{\lambda_0}*f_1$, where $e_{\lambda_0}(t)=\lambda_0^t e^{-\lambda_0 t}$, $t\geq 0$, and f_1 is the pdf of the first explosion time starting from $i_0=1$. Check that $\sum_{n=1}^{\infty}f_0^{*n}(t)=\lambda_0\tilde{p}_{10}(t)$.]

7. Compute the transition probabilities of the homogeneous Poisson process by Feller's method of successive approximation.
8. *(A Pure Death Process)* Let $S = \{0, 1, 2, \ldots\}$ and let $q_{i,i-1} = \delta > 0$, $q_{ii} = -\delta$, $i \geq 1$, $q_{ij} = 0$, otherwise.

   (i) Calculate $p_{ij}(t) \equiv p(t; i, \{j\}))$, $t \geq 0$ using successive approximations.
   (ii) Calculate $\mathbb{E}_i X_t$ and $\mathrm{Var}_i X_t$.

9. *(Birth–Death with Linear Rates)* The continuous parameter birth–death chain with *linear rates* on $S = \{0, 1, 2, \ldots\}$ is defined by the (minimal) Markov chain $\{X_t : t \geq 0\}$ with $q_{i,i+1} = i\beta$, $q_{i,i-1} = i\delta$, $i \geq 0$, where $\beta, \delta > 0$, $q_{ij} = 0$ if $|j - i| > 1$, $q_{01} = 0$. Let

$$m_i(t) = \mathbb{E}_i X_t, \qquad s_i(t) = \mathbb{E}_i X_t^2.$$

   (i) Use the forward equation to show $m_i'(t) = (\beta - \delta)m_i(t)$, $m_i(0) = i$.
   (ii) Show $m_i(t) = ie^{(\beta-\delta)t}$.
   (iii) Show $s'(t) = 2(\beta - \delta)s_i(t) + (\beta + \delta)m_i(t)$.
   (iv) Show that

$$s_i(t) = \begin{cases} ie^{2(\beta-\delta)t}\{i + \frac{\beta+\delta}{\beta-\delta}(1 - e^{-(\beta-\delta)t})\}, & \text{if } \beta \neq \delta \\ i(i + 2\beta t), & \text{if } \beta = \delta. \end{cases}$$

   (v) Calculate $\mathrm{Var}_i X_t$.

10. *(Pure Birth with Affine Linear Rates)* Consider a pure birth process on $S = \{0, 1, 2, \ldots\}$ having infinitesimal parameters $q_{ii} = -(\nu + i\lambda)$, $q_{i,i+1} = \nu + i\lambda$, $q_{ik} = 0$ if $k < i$ or if $k > i + 1$, where $\nu, \lambda > 0$. Successively solve the forward equation for $p_{00}(t)$, $p_{01}(t)$ and inductively for $p_{0n}(t) = \frac{\nu(\nu+\lambda)\cdots(\nu+(n-1)\lambda)}{n!\lambda^n} e^{-\nu t}[1 - e^{-\lambda t}]^n$. The case $\nu = 0$ is referred to as the *Yule linear growth process* (see Exercise 3, Chapter 1).
11. Show that a discrete parameter birth and death process on $S = \{0, 1, \ldots\}$ with reflecting boundary at 0 and $\beta_j \delta_j > 0$, $j \geq 1$, is transient if and only if $\sum_{n=1}^{\infty} \frac{\delta_1 \cdots \delta_j}{\beta_1 \cdots \beta_j} < \infty$. [*Hint*: Compute the probability of hitting i before k from an initial state $j \in i, \ldots, k$, and the probability of eventually hitting i by passing to the limit as $k \to \infty$.]
12. Consider a recurrent birth and death process on $S = \{0, 1, \ldots\}$ with reflecting boundary at 0 and $\beta_j \delta_j > 0$, $j \geq 1$. (a) Show $\sum_{j=1}^{\infty} \frac{\beta_0 \cdots \beta_{j-1}}{\delta_1 \cdots \delta_j} < \infty$ implies, for suitable choice of normalizing constant m_0, $m_j = \frac{\beta_0 \cdots \beta_{j-1}}{\delta_1 \cdots \delta_j} m_0$ is an invariant probability for the embedded discrete parameter Markov chain. [*Hint*: Check time-reversibility.] (b) Define $\pi_1 = \frac{\lambda_0}{\lambda_1 \delta_1} \pi_0$, $\pi_j = \frac{\lambda_0}{\lambda_j} m_j$, $j \geq 2$. Show that if $\sum_{j=1}^{\infty} \frac{\lambda_0}{\lambda_j} m_j < \infty$, then, for suitable choice of π_0, $\{\pi_j\}_{j=0}^{\infty}$ is the unique

invariant probability for the continuous parameter Markov chain. [*Hint*: Check that the forward equations apply.]

13. *(α-Riccati Jump Process)* The α-Riccati process arises in a variety of applications[3] and can be associated with a variety of critical phenomena. Let $\mathbb{T} = \cup_{n=0}^{\infty}\{1,2\}^n, \quad \{1,2\}^0 = \{\emptyset\}$, and view $\mathbb{T}$ as a *binary tree graph* rooted at $\emptyset$. The *height* of $v = (v_1, \ldots, v_n) \in \mathbb{T}$ is defined by $|v| = n$ and $|\emptyset| = 0$. If $v \in \mathbb{T}$, then $vj = (v_1, \ldots, v_n, j)$. $\emptyset j = (j)$, $j \in \{1, 2\}$ denotes the first-generation offspring of v, and $v|m = (v_1, \ldots, v_m), m \le n = |v|$ is the m-generation ancestor of v. For $|v| = n$, $\overleftarrow{v} := v|(n-1)$ denotes the immediate ancestor of v. Then, $u, v \in \mathbb{T}$ are connected by an *edge* if $u = \overleftarrow{v}$, or $v = \overleftarrow{u}$ to make $\mathbb{T}$ a *graph*. The α-Riccati process is the jump Markov process on the denumerable *evolutionary space* $S \equiv \mathcal{E}$ consisting of finite subsets $V \subset \cup_{n=0}^{\infty}\{1,2\}^n$ with the properties that are as follows: (i) $V = \{\emptyset\} \in \mathcal{E}$. (ii) If $\{\emptyset\} \ne V \in \mathcal{E}$, then there is a $v \in V$ such that $V^{(v)} = V\backslash\{v\} \cup \{v1, v2\} \in \mathcal{E}$. Define $A^{(\alpha)}f(V) = \sum_{v\in V} \alpha^{|v|}\{f(V^v) - f(V)\} = \lambda_\alpha(V) \sum_{W\in\mathcal{E}}\{f(W) - f(V)\}k_\alpha(V, W),\ V \in \mathcal{E}$, where $\lambda_\alpha(V) = \sum_{v\in V}\alpha^{|v|},\ V \in \mathcal{E}$ and

$$k_\alpha(V, W) = \begin{cases} \alpha^{|v|}/\lambda_\alpha(V), & \text{if } W = V^v \text{for some } v \in V, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Show[4] that the $\alpha = 1/2$-Riccati is a standard Poisson process. (b) Show that the α-Riccati is explosive if and only if $\alpha > 1$. [*Hint*: Let $\lambda_\beta(V) = \sum_{v\in V}\beta^{|v|}$ and observe that $A^{(\alpha)}\lambda_\beta = (2\beta - 1)\lambda_{\alpha\beta}$. Also note that $\lambda_1(V) = |V|$ is the cardinality of V. To prove explosion, consider $u(V) = \frac{\sqrt{|V|}}{\sqrt{|V|+1}}$, $V \in \mathcal{E}$. Use $\sqrt{|V|+1} + 1 \le 2\sqrt{|V|+1}$ and $\lambda_\alpha(V) \ge |V|$ for $\alpha > 1$, to check $A^{(\alpha)}u(V) \ge \{\frac{\sqrt{|V|+1}-\sqrt{|V|}}{2\sqrt{|V|+1}}\sqrt{|V|}\}u(V) \quad V \in \mathcal{E}$, and the function $x \to \sqrt{x}\frac{\sqrt{x+1}-\sqrt{x}}{2\sqrt{1+x}}$, $x \ge 1$, is bounded below by $\gamma = \frac{\sqrt{2}-1}{2\sqrt{2}} < 1$. To prove nonexplosion, consider $u(V) = |V|$.]

14. Consider a pure birth process on $S = \{0, 1, 2, \ldots\}$ defined by infinitesimal parameters $q_{k,k+1} = \lambda_k > 0$, $q_{kk} = -\lambda_k$.
    (i) If $\sum_0^\infty \lambda_k^{-1} = \infty$ and $\sum_0^\infty \lambda_k^{-2} < \infty$, then show that the variance of $\tau_n = T_0 + T_1 + \cdots + T_{n-1}$ (the time to reach n, starting from 0) goes to a finite limit as $n \to \infty$. Use Kolmogorov's zero-one law to show that as $n \to \infty$, $\tau_n - \sum_0^{n-1}\lambda_k^{-1}$ converges (for all sample paths, outside a set of probability zero) to a finite random variable η.
    (ii) If $\sum_0^\infty \lambda_k^{-1} = \infty = \sum_0^\infty \lambda_k^{-2}$ but $\sum_0^\infty \lambda_k^{-3} < \infty$, show that

[3] See Athreya (1985), Aldous and Shields (1988), Best and Pfaffelhuber (2010), Bhattacharya and Waymire (2021), and Dascaliuc et al. (2018a,b, 2023a,b,c). This includes a role for α-Riccati as an idealized model for self-similar incompressible Navier-Stokes equations.

[4] See Aldous and Shields (1988) and Dascaliuc et al. (2018a).

$$(\tau_n - \sum_0^{n-1} \lambda_k^{-1})/(\sum_0^{n-1} \lambda_k^{-2})^{1/2}$$

is asymptotically (as $n \to \infty$) Gaussian with mean 0 and variance 1. [*Hint*: Use a central limit theorem, BCPT,[5] p. 81.]

15. Let $\{X_t : t \geq 0\}$ be a Poisson process with parameter $\lambda > 0$. Let T_0 denote the time of the first occurence.
    (i) Let $N = X_{2T_0} - X_{T_0}$ and calculate $\text{Cov}(N, T_0)$.
    (ii) Calculate the (conditional) expected value of the time $T_0 + \cdots + T_{r-1}$ of the rth occurence given $X_t = n > r$.
16. Prove the transience and recurrence dichotomy of a state j for a discrete parameter Markov chain. [*Hint*: Use the strong Markov property.]
17. *(Continuous Parameter Branching Markov Chain)* At time 0, $X_0 = i_0 \geq 1$. Each of the i_0 particles, independently of the others, waits an exponential length of time with parameter $\lambda > 0$ and then splits into a random number k particles with probability f_k, independently of the holding time, where $f_k \geq 0$, $f_0 > 0$, $0 < f_0 + f_1 < 1$, and $\sum_{k=0}^{\infty} f_k = 1$. The progeny continues to split according to this process, and X_t counts the total number present at time $t \geq 0$. If $X_s = 0$, then $X_t = 0$ for all $t \geq s$, i.e., 0 is an *absorbing* state. Determine the infinitesimal parameters $Q = ((q_{ij})), \textit{and}\ \lambda_i, k_{ij} \equiv k(i, \{j\})$.
18. Let $p(t) = ((p_{ij}(t)))$ denote the transition probabilities for the continuous parameter branching Markov process given in Exercise 17. Let $g_t^{(i)}(r) = \sum_{j=0}^{\infty} p_{ij}(t) r^j$, $i = 0, 1, 2, \ldots$.
    (i) Use the Chapman–Kolmogorov equations to show $g_{t+s}^{(i)}(r) = g_t^{(i)}(g_s^{(1)}(r))$.
    (ii) Let $\rho = P(X_t = 0 \text{ for some } t > 0 | X_0 = 1)$. Show that ρ is the smallest nonnegative root of the equation $h(r) = 0$, where $h(r) = \sum_{j=0}^{\infty} q_{1j}^{j}$.
    (iii) Show that if $\sum_{j=0}^{\infty} j f_j \leq 1$, then $\rho = 1$.
19. *(Continuous Parameter Critical Binary Branching)* Consider the critical binary continuous parameter branching Markov chain[6] in $\{X_t : t \geq 0\}$ defined by the offspring distribution $f_0 = f_1 = \frac{1}{2}$ in the context of Exercise 17.
    (i) Verify that $\{X_t : t \geq 0\}$ is a *birth–death process* on $S = \{0, 1, 2, \ldots\}$ with absorption at 0 and birth–death rates $q_{ii+1} = q_{ii-1} = i\lambda/2$, $i \geq 1$.
    (ii) Let $M_t = \#\{0 \leq s \leq t : X_s - X_{s-} = -1\}$ denote the number of deaths in time 0 to t. Show that $\{(X_t, M_t) : t \geq 0\}$ is a continuous parameter Markov chain on $S = \mathbb{Z}^+ \times \mathbb{Z}^+$ with infinitesimal rates $A = ((a_{\mathbf{ij}}))$,

[5] Throughout, BCPT refers to Bhattacharya and Waymire (2016), A Basic Course in Probability Theory.

[6] This exercise is based on questions about the main channel lengths in a river network by Gupta et al. (1990). Also see Durrett et al. (1991).

where $a_{\mathbf{i},\mathbf{j}} = \frac{\lambda}{2} i_1$ if $\mathbf{j} = \mathbf{i} + (1, 0)$ or $\mathbf{j} = \mathbf{i} + (-1, 1)$, $a_{\mathbf{ij}} = -\lambda i_1$ if $\mathbf{i} = \mathbf{j}$, and $a_{\mathbf{i},\mathbf{j}} = 0$ otherwise.

(iii) Let $g(t, u, v) = \mathbb{E}_{(1,0)} u^{X_t} v^{M_t}$, $0 \le u, v \le 1$, $t \ge 0$. Use Kolmogorov's forward equation to show $\frac{\partial g}{\partial t} = \frac{\lambda}{2}(u^2 + v - 2u)\frac{\partial g}{\partial u}$, $g(0, u, v) = v$.

(iv) Integrate the characteristic curve defined by

$$\gamma'(t) = -\frac{\lambda}{2}\, j(\gamma^2(t) - 2\gamma(t)tv), \quad \frac{d}{dt}\, g(t, \gamma(t), v) = 0,$$

$g(0, \gamma(0), v) = \gamma(0)$ to obtain $g(t, u, v) = \frac{c_1(u-c_2)+c_2(c_1-u)e^{\frac{1}{2}\lambda(c_1-c_2)t}}{(u-c_2)-(u-c_1)e^{\frac{1}{2}\lambda(c_1-c_2)t}}$ where $c_1 = c_1(v) = 1 - (1 - v)^{V_2}$, $c_2 = c_2(v) = 1 + (1 - v)^{\frac{1}{2}}$.

(v) Let $\tau_0 = \inf\{t \ge 0 : X_t = 0\}$. Show that

$$P\left(M_{\tau_0} = n\right) = \binom{\frac{1}{2}}{n}(-1)^{n-1} = \frac{1}{2n-1}\binom{2n-1}{n}2^{-2n+1}, \; n = 1, 2, \ldots$$

[*Hint*: $\sum_{n=1}^{\infty} P(M_{\tau_0} = n) = \lim_{v\to\infty} g(t, 1, v)$, $0 \le v < 1$.]

(vi) Using Stirling's formula show that

$$P\left(M_{\tau_0} = n\right) \sim \frac{1}{2\sqrt{\pi}}\, n^{-\frac{3}{2}} \quad \text{as } n \to \infty.$$

(vii) Let $h_n = P(M_{\tau_0} = n)\mathbb{E}\{\tau_0 \mid M_{\tau_0} = n\}$, $n \ge 1$, and $\hat{h}(v) = \sum_{n=1}^{\infty} h_n v^n$. Show that

$$\hat{h}(v) = -\frac{2}{\lambda}\log(1 - \frac{c_1(v)}{c_2(v)}) = -\frac{2}{\lambda}\log(\frac{2\sqrt{1-v}}{1+\sqrt{1-v}}).$$

[*Hint*: $P(\tau_0 > t \mid M_{\tau_0} = n) = \frac{P(M_{\tau_0}=n)-P(X_t=0, M_{\tau_0}=n)}{P(M_{\tau_0}=n)}$.]

(viii) $\mathbb{E}\{\tau_0 \mid M_{\tau_0} = n\} \sim \frac{2\sqrt{\pi}}{\lambda}\, n^{\frac{1}{2}}$ as $n \to \infty$. [*Hint*: Apply a Tauberian theorem for Laplace transforms, Feller 1971, Chapter XIII.6.]

Chapter 5
Processes with Independent Increments

The goal is to characterize all processes continuous in probability having independent increments and certain sample path regularity as a sum of (possibly degenerate) continuous Brownian motion and a limit of independent Poisson superpositions of terms representing jumps. It should be noted that while we restrict to $\mathbb{R}$-valued processes for simplicity of exposition, the methods and results readily extend to multidimensional $\mathbb{R}^k$-valued processes.

Let us begin by revisiting some essential examples of processes with independent increments from the perspective of the present chapter.[1]

Example 1 (Poisson and Compound Poisson Processes) Let $N = \{N_t : t \geq 0\}$ denote a Poisson process with intensity parameter $\rho > 0$ and $Y_0 = 0, Y_1, Y_2, \ldots$ an i.i.d. sequence of nonnegative random variables distributed as $\gamma(dy)$, independently of N. Consider the compound Poisson process defined by

$$X_t = \sum_{j=1}^{N_t} Y_j, \; X_0 = 0, \quad t > 0.$$

[1] The primary focus of this chapter is the representation theory of stochastic processes having stationary, independent increments. A variety of interesting applications can be found in Woyczyński (2001) and in Barndorff-Nielsen et al. (2001). The approach here essentially follows that of Itô and Rao (1961), Itô (2004).

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_5

Let

$$\psi(\lambda) = \mathbb{E}e^{-\lambda Y_1}, \quad \lambda \geq 0$$

The assumption of nonnegative summands assures finiteness of $\psi(\lambda)$. For the general theory, Laplace transforms are replaced by characteristic functions. It is straightforward to check that $X = \{X_t : t \geq 0\}$ is a right-continuous stochastic process with stationary and independent increments. In fact, the sample paths are right-continuous step functions. Then, conditioning on N, one has

$$\begin{aligned}
\mathbb{E}e^{-\lambda(X_{t+s}-X_s)} &= \mathbb{E}e^{-\lambda X_t} = \mathbb{E}(\psi(\lambda))^{N_t} \\
&= \sum_{n=0}^{\infty} \frac{(\rho t \psi(\lambda))^n}{n!} e^{-\rho t} = \exp\{\rho t(\psi(\lambda) - 1)\} \\
&= \exp\{-t \int_{\mathbb{R}} (1 - e^{-\lambda y}) \nu(dy)\}, \quad \lambda \geq 0, \qquad (5.1)
\end{aligned}$$

where

$$\nu(dy) = \rho\gamma(dy).$$

Example 2 (Lévy's Nondecreasing Processes with Stationary Independent Increments) In this example, we consider a structural lemma that includes the Poisson process and the compound Poisson process for nonnegative Y_1 but is significantly richer than these examples.

Lemma 1 (Lévy (1954)) *Suppose that X is a right-continuous process having stationary independent increments such that*

$$0 \leq X_0 \leq X_s \leq X_t, \quad 0 \leq s \leq t.$$

Then,

$$\mathbb{E}e^{-\lambda(X_t - X_0)} = e^{-t\psi(\lambda)}, \quad \lambda > 0,$$

where

$$\psi(\lambda) = m\lambda + \int_0^{\infty} (1 - e^{-\lambda x}) \nu(dx),$$

for a constant $m \geq 0$ and a Borel measure $\nu(dx)$ on $(0, \infty)$, such that $\int_0^{\infty}(1 - e^{-x})\nu(dx) < \infty$.

Proof Without loss of generality assume that $X_0 = 0$, else replace X_t by $X_t - X_0$ for $t \geq 0$. Observe from stationary independent increments, writing $g(t) = \mathbb{E}e^{-\lambda X_t}$ for fixed $\lambda > 0$, one has

$$g(t) = g(t-s)g(s), \quad 0 \le s < t$$

and $0 < g(t) \le 1, t \ge 0$. Thus, $g(t)$ is an exponential function of t with nonnegative parameter implicitly depending on λ, say, (see Exercise 1, Chapter 4)

$$\mathbb{E}e^{-\lambda X_t} = g(t) = e^{-t\psi(\lambda)}, \quad t \ge 0,\ 0 \le \psi(\lambda) < \infty. \tag{5.2}$$

Noting that $x \to xe^{-\lambda x}$, $x \ge 0$, is a bounded function for $\lambda > 0$ and differentiating (5.2) with respect to $\lambda > 0$, one arrives at

$$\psi'(\lambda)e^{t\psi(\lambda)} = \frac{1}{t}\int_0^\infty xe^{-\lambda x}P(X_t \in dx), \quad t > 0, \lambda > 0.$$

Since the left side has a limit, namely, $\psi'(\lambda)$, as $t \downarrow 0$, the same must be true on the right side. Notice that for each $t > 0$, the right side is the Laplace transform of a measure $\frac{x}{t}P(X_t \in dx)$. Thus, in view of convergence on the left side, the limit a-priori exists as $t \downarrow 0$ on the right side and is, therefore, also a Laplace transform of a measure, say $\tilde{\nu}(dx)$, by the extended continuity theorem[2] for Laplace transforms. That is,

$$\psi'(\lambda) = \int_0^\infty e^{-\lambda x}\tilde{\nu}(dx), \quad \lambda > 0. \tag{5.3}$$

Also, $\psi(0) = 0$. Hence, integrating both sides of (5.3), one obtains using the Fubini-Tonelli theorem,

$$\begin{aligned}
\psi(\lambda) &= \int_0^\lambda \int_0^\infty \exp\{-yx\}\tilde{\nu}(dx)dy \\
&= \int_0^\lambda \{\int_{\{0\}} \exp\{-yx\}\tilde{\nu}(dx) + \int_{0+}^\infty \exp\{-yx\}\tilde{\nu}(dx)\}dy \\
&= \lambda\tilde{\nu}(\{0\}) + \int_{0+}^\infty \frac{1-e^{-\lambda x}}{x}\tilde{\nu}(dx) \\
&= \lambda m + \int_{0+}^\infty (1-e^{-\lambda x})\nu(dx), \qquad (5.4)
\end{aligned}$$

where $m = \int_{\{0\}} \exp\{-yx\}\tilde{\nu}(dx) = \tilde{\nu}(\{0\})$ and $\nu(dx) = \frac{1}{x}\tilde{\nu}(dx)$ on $(0, \infty)$. Since $\int_{0+}^\infty (1-e^{-x})\nu(dx) \le \psi(1) < \infty$, the proof of the lemma is now complete. ■

---

[2] See Feller (1968, 1971), p. 433.

Remark 5.1 Stochastic processes of the type considered in Lemma 1 are often referred to as *subordinators* for their role as "random clocks" subordinate to other processes (see Example 3).

Lévy in fact went much further than this lemma, to derive the following sample path representation in terms of a Poisson random measure $N(dt \times dx)$ with *intensity measure* $dt \times \nu(dx)$. In particular,

$$P(N(A) = n) = \frac{(\int_A dt\nu(dx))^n}{n!} e^{-\int_A dt\nu(dx)}, \ A \in \mathcal{B}([0,\infty) \times (0,\infty)), n = 0, 1 \dots.$$

For ease of reference, the complete definition of the *Poisson random measure (or Poisson random field)* is as follows.

Definition 5.1 A *Poisson random measure* on a measurable space $(S, \mathcal{S})$ is a random field $N = \{N(A) : A \in \mathcal{S}\}$, i.e., a stochastic process indexed by $\mathcal{S}$, defined on a probability space $(\Omega, \mathcal{F}, P)$ such that

a. $N(A)$ has a Poisson distribution for each $A \in \mathcal{S}$.
b. $N(A_1), \dots, N(A_n)$ are independent for any disjoint $A_1, \dots, A_n, n \geq 1$ in $\mathcal{S}$.
c. For almost all $\omega \in \Omega$, $A \to N(A, \omega)$, $A \in \mathcal{S}$, is a measure.

$\rho(A) = \mathbb{E}N(A)$, $A \in \mathcal{S}$, is referred to as the intensity measure.

That the set function $A \to \rho(A), A \in \mathcal{S}$, is a measure follows immediately from Lebesgue's monotone convergence theorem. The following lemma provides a formula for the so-called *Laplace functional* of the Poisson random field.

Proposition 5.1 *Suppose that N is a Poisson random field on $(S, \mathcal{S})$ with intensity measure $\rho(dz)$. If f is a non-negative measurable function on S, then*

$$\mathbb{E}e^{-\lambda \int_S f(z)N(dz)} = \exp\{\int_S (1 - e^{-\lambda f(z)})\rho(dz)\}, \quad \lambda \geq 0. \tag{5.5}$$

Proof Suppose that $f(z) = a\mathbf{1}_A(z)$, $z \in S$, for a constant $a \geq 0$ and measurable set $A \in \mathcal{S}$. Then,

$$\begin{aligned}\mathbb{E}e^{-\lambda \int_S f(z)N(dz)} &= \mathbb{E}e^{-\lambda a N(A)} \\ &= \exp\{-\rho(A)(1 - e^{-\lambda a})\} \\ &= \exp\{\int_S (1 - e^{-\lambda f(z)})\rho(dz)\}, \end{aligned} \tag{5.6}$$

since $1 - e^{-\lambda f(z)} = (1 - e^{-\lambda a})\mathbf{1}_A(z)$. The formula extends to $f(z) = \sum_{j=1}^m a_j \mathbf{1}_{A_j}(z)$, $z \in S$, for $a_j \geq 0$, and disjoint measurable $A_j \in \mathcal{S}$, $j = 1, \dots, m$, by independence of "increments," since

$$\begin{aligned}
\mathbb{E}e^{-\lambda \int_S f(z)N(dz)} &= \mathbb{E}\prod_{j=1}^{m} e^{-\lambda a_j N(A_j)} \\
&= \prod_{j=1}^{m} \exp\{-\rho(A_j)(1-e^{-\lambda a_j})\} \\
&= \exp\{-\sum_{j=1}^{m}\rho(A_j)(1-e^{-\lambda a_j})\} \\
&= \exp\{\int_S (1-e^{-\lambda f(z)})\rho(dz)\}. \qquad (5.7)
\end{aligned}$$

From here, one may use Lebesgue's monotone convergence theorem to obtain the general formula (Exercise 2). ■

Corollary 5.2 *The stochastic process given in Lemma 1 is stochastically equivalent to the process defined by*

$$X_t = X_0 + mt + \int_0^t \int_{0+}^{\infty} xN(ds \times dx), \quad t \ge 0, \qquad (5.8)$$

where N is the Poisson random field on $S = [0, \infty) \times [0, \infty)$ with intensity measure $\rho(dt \times dx) = dt\nu(dx)$.

Proof Without loss of generality, assume $X_0 = 0$, or replace X_t by $X_t - X_0, t \ge 0$ in the following. By the corresponding properties of the Poisson random measure N, (5.8) defines a process with stationary and independent increments. To also see that this process is equivalent in distribution to the process of Lemma 1, it suffices, by the uniqueness theorem for Laplace transforms[3] to calculate $\mathbb{E}e^{-\lambda X_t}$ for fixed $t > 0$. Let $f_t(s, x) = x\mathbf{1}_{[0,t]}(s), s, x \ge 0$ Then, applying (5.5) to the function f_t, one has

$$\begin{aligned}
\mathbb{E}e^{-\lambda X_t} &= e^{-\lambda mt}\mathbb{E}e^{-\lambda \int_0^t \int_{0+}^{\infty} xN(ds\times dx)} \\
&= e^{-\lambda mt}\mathbb{E}e^{-\lambda \int_{[0,\infty)\times[0,\infty)} f_t(s,x)N(ds\times dx)} \\
&= \exp\{-tm\lambda - t\int_{0+}^{\infty}(1-e^{-\lambda x})\nu(dx)\}. \qquad (5.9)
\end{aligned}$$

This calculation, taken together with the stationary and independent increments property, proves stochastic equivalence between the process analyzed in Lemma 1 and the process defined by (5.8). ■

[3] See Feller (1968, 1971), Chapter XIII.

Remark 5.2 Lemma 1 and Corollary 5.2 characterize all processes with stationary, nonnegative independent increments having cádlág sample paths. Adding a term ct is redundant if $c > 0$, but if $c < 0$, such an addition provides a slightly enhanced class of processes.

An additional consequence for the structure of the Laplace transform occurs if, in addition to the conditions of Lemma 1, one further assumes and a *scaling relation* of the form (see Example 3): For each $\lambda > 0$, there is a $\alpha = \alpha(\lambda) > 0$, such that

$$X_t =^{dist} \lambda X_{\alpha^{-1}t}, \quad t \geq 0, X_0 = 0, \tag{5.10}$$

or, equivalently, $\lambda X_t =^{dist} X_{\alpha^{-1}(\frac{1}{\lambda})t}, \lambda > 0, t \geq 0$. In particular, it follows, by iterating the scaling relation (Exercise 4), that one has

$$\alpha(\lambda_1\lambda_2) = \alpha(\lambda_1)\alpha(\lambda_2), \quad \lambda_1, \lambda_2 > 0. \tag{5.11}$$

Thus, one may apply Cauchy's functional formula to $\alpha(e^x), x \in \mathbb{R}$, (see Exercise 1 in Chapter 4), to conclude that

$$\alpha \equiv \alpha(\lambda) = \lambda^\theta, \quad \text{for some } \theta \geq 0. \tag{5.12}$$

Such processes are referred to as one-sided stable processes , and the exponent θ is referred to as a *stable law scaling exponent*. The case in which $\theta = 0$ corresponds to the degenerate case, in which $X_t = 0$ a.s.

Corollary 5.3 *In addition to the conditions of Lemma 1, assume the scaling relation (5.10), with scaling exponent* θ*. Then,* $0 \leq \theta \leq 1$*. Moreover,*

$$X_t = \begin{cases} \psi(1)t & \text{if } \theta = 1, \\ mt + \int_0^t \int_{0+}^\infty xN(ds \times dx), & \text{if } 0 \leq \theta < 1, \end{cases}$$

where for $\theta < 1$*,*

$$\nu(dx) = \frac{\psi(1)\theta}{\Gamma(1-\theta)x^{\theta+1}}dx,$$

and $N(ds \times dx)$ *is the Poisson random measure on* $(0, \infty) \times (0, \infty)$ *with intensity measure* $dt \times \nu(dx)$ *and* $m \geq 0$*.*

Proof Observe for arbitrary $t, \lambda \geq 0$, one has

$$e^{-t\psi(\lambda)} = \mathbb{E}e^{-\lambda X_t} = \mathbb{E}e^{-X_{\lambda^\theta t}} = e^{-\lambda^\theta t\psi(1)}.$$

Thus,

$$\psi(\lambda) = \psi(1)\lambda^\theta, \quad \lambda \geq 0.$$

Using Lemma 1,

$$\psi(1)\lambda^\theta = \psi(\lambda) = \mu\lambda + \int_{0+}^\infty (1 - e^{-\lambda x})\nu(dx).$$

In particular, $\theta = 1$ corresponds to a degeneracy for which $X_t = \gamma t$, for some γ, and $\psi(1) = \log \mathbb{E}e^{X_1} = \gamma$. For $\theta > 0$,

$$\frac{d}{d\lambda}\int_{0+}^\infty (1 - e^{-\lambda x})\nu(dx) = \frac{d}{d\lambda}\psi(1)\lambda^\theta,$$

yields that the Laplace transform of $\tilde{\nu}(dx) = x\nu(dx)$ is given by $\psi(1)\theta\lambda^{\theta-1}$, $\lambda > 0$. Since the Laplace transform is a non-increasing function of $\lambda > 0$, one has $\theta < 1$. From here, one may identify the asserted formula $\tilde{\nu}(dx) = x\nu(dx) = \frac{\psi(1)\theta}{\Gamma(1-\theta)x^\theta}dx$ as the inverse Laplace transform of $\psi(1)\theta\lambda^{\theta-1}$, $\lambda > 0$, by using a simple change of variables ($y = \lambda x$) as follows:

$$\int_{0+}^\infty e^{-\lambda x}\frac{\psi(1)}{\Gamma(1-\theta)x^\theta}dx = (\frac{\psi(1)}{\Gamma(1-\theta)}\int_{0+}^\infty y^{-\theta}e^{-y}dy)\lambda^{\theta-1} = \psi(1)\lambda^{\theta-1}.$$

■

Example 3 (First Passage Time Processes) [4] Let $\{B_t\}$ be a standard Brownian motion starting at zero. Fix $0 < a < \infty$. The first passage time to a is defined by

$$\tau_a := \inf\{t \geq 0 : B_t \geq a\}, \quad a \in \mathbb{R}.$$

Consider the nondecreasing stochastic process $\{\tau_a : a \geq 0\}$.

Proposition 5.4 *Under P_0, the stochastic process $\{\tau_a : 0 \leq a < \infty\}$ is a process with independent increments. Moreover, the increments are homogeneous, i.e., for $0 < a < b, 0 < c < d$, the distribution of $\tau_b - \tau_a$ is the same as that of $\tau_d - \tau_c$ if $d - c = b - a$, both distributions being the same as the distribution of τ_{b-a} under P_0.*

Proof Recall that the conditional distribution of the *after-τ_a* process $B^+_{\tau_a}$ given the past up to time τ_a is $P_{B_{\tau_a}} = P_a$, since $B_{\tau_a} = a$ on the event $[\tau_a < \infty]$ and $P_0(\tau_a < \infty) = 1$. Thus, $\{B_{t\wedge\tau_a}\}$ and $\{B_{\tau_a+t}\}$ are *independent stochastic process* and the distribution of $B^+_{\tau_a}$ under P_0 is P_a. That is, $B^+_{\tau_a} := \{(B^+_{\tau_a})_t\}$ has the same distribution as that of a Brownian motion starting at a. Also, starting from zero, the state b can be reached only after a has been reached. Hence, with P_0-probability 1,

[4] This example partially motivated the introduction of the *inverse Gaussian* distributions by Barndorff-Nielsen and Halgreen (1977) and Barndorff-Nielsen et al. (1978).

$$\tau_b = \tau_a + \inf\{t \geq 0 : X_{\tau_a+t} = b\} = \tau_a + \tau_b\left(X^+_{\tau_a}\right), \tag{5.13}$$

where $\tau_b(B^+_{\tau_a})$ is the first hitting time of b by the after-τ_a process $B^+_{\tau_a}$. Since this last hitting time depends only on the after-τ_a process, it is independent of τ_a, which is measurable with respect to $\mathcal{F}_{\tau_a}$. Hence, $\tau_b - \tau_a \equiv \tau_b(B^+_{\tau_a})$ and τ_a are independent, and the distribution of $\tau_b - \tau_a$ under P_0 is the distribution of τ_b under P_a. This last distribution is the same as the distribution of τ_{b-a} under P_0. For a Brownian motion $\{B_t\}$ starting at a, $\{B_t - a\}$ is a Brownian motion starting at zero; and if τ_b is the first passage time of $\{B_t\}$ to b, then it is also the first passage time of $\{B_t - a\}$ to $b - a$. Finally, if $0 < a_1 < a_2 < \cdots < a_n$ are arbitrary, then these arguments applied to $\tau_{a_{n-1}}$ shows that $\tau_{a_n} - \tau_{a_{n-1}}$ is independent of $\{\tau_{a_i} : 1 \leq i \leq n-1\}$, and its distribution (under P_0) is the same as the distribution of $\tau_{a_n - a_{n-1}}$ under P_0. ■

Thus, the first passage process for Brownian motion starting at zero falls within the framework of Example 2, with the stochastically equivalent representation presented there. In particular, also note that $\tau_0 = 0$ and, for $\lambda > 0$,

$$\begin{aligned}\lambda\tau_a &= \inf\{\lambda t : B_t \geq a\}\\ &= \inf\{\lambda t : \lambda^{-\frac{1}{2}} B_{\lambda t} \geq a\} = \tau_{\lambda^{\frac{1}{2}} a}.\end{aligned} \tag{5.14}$$

Thus,

$$\alpha(\lambda) = \lambda^{\frac{1}{2}}, \quad \theta = 1/2. \tag{5.15}$$

For contrast, let us now consider a (non-degenerate) process having stationary independent increments and almost sure continuous sample paths.

Example 4 (Brownian Motion) The Brownian motion starting at zero, $X = \{X_t = \mu t + \sigma B_t : t \geq 0\}, X_0 = 0$, is a stochastic process having continuous sample paths with stationary and independent Gaussian increments, uniquely specified by the parameters $\mu, \sigma^2 > 0$. Moreover,

$$\mathbb{E}e^{-\lambda X_t} = e^{-\lambda\mu t + \frac{1}{2}\lambda^2\sigma^2 t}, \quad \lambda \in \mathbb{R}. \tag{5.16}$$

Almost sure sample path continuity uniquely sets this example apart from the previous (non-degernate) examples, thereby occupying a special place in the general theory.

In the spirit of these examples, the goal of this chapter is to characterize all processes continuous in probability and having independent increments and certain sample path regularity as a sum of (possibly degenerate) continuous Brownian motion and a limit of independent Poisson superpositions of terms representing

jumps. The founders[5] of the theory are Bruno de Finetti, Paul Lévy, Yu Khinchine, and A.N. Kolmogorov. Our exposition follows that of K. Itô.

Definition 5.2

a. A stochastic process $X = \{X_t : t \geq 0\}$, $X_0 = 0$, having independent increments is referred to as an *additive process*.
b. A *Lévy process* is an additive stochastic process $\{X_t : t \geq 0\}$, continuous in probability, whose sample paths are a.s. right-continuous with left-hand limits.
c. A Lévy process having stationary increments will be referred to as a *homogeneous Lévy process*.

Remark 5.3 The condition that $X_0 = 0$ is by standard convention. Naturally, all else being satisfied, one may always replace X_t by $X_t - X_0$ if needed.

Brownian motion and the compound Poisson process are prototypical examples of homogeneous Lévy processes. This is explained by the following two theorems and their corollaries.

Theorem 5.5 *If* $\{X_t : t \geq 0\}$ *is an additive process having a.s. continuous sample paths, then the increments of* $\{X_t : t \geq 0\}$ *must be Gaussian.*

Proof Let $s < t$. Recall that a continuous function on a compact interval is uniformly continuous. Thus,

$$P(\cap_{n=1}^{\infty} \cup_{k=1}^{\infty} [|X_u - X_v| < \frac{1}{n}, \text{whenever } |u - v| < \frac{1}{k}, s \leq u \leq t]) = 1.$$

So, by sample path continuity, for any $\varepsilon > 0$, there is a number $\delta = \delta(\varepsilon) > 0$ such that

$$P(|X_u - X_v| < \varepsilon \text{ whenever } |u - v| < \delta, s \leq u, v \leq t) > 1 - \varepsilon.$$

Let $\{\varepsilon_n\}_{n=1}^{\infty}$ be a sequence of positive numbers decreasing to zero and partition $(s, t]$ into subintervals $s = t_0^{(n)} < t_1^{(n)} < \cdots < t_{k_n}^{(n)} = t$ of lengths less than $\delta_n = \delta(\varepsilon_n)$. Then,

$$X_t - X_s = \sum_{j=1}^{k_n} \{X(t_j^{(n)}) - X(t_{j-1}^{(n)})\} := \sum_{j=1}^{k_n} X_j^{(n)}$$

where $X_1^{(n)}, \ldots, X_{k_n}^{(n)}$ are independent (*triangular array*). Observe that the truncated random variables

[5] See Bose et al. (2002) for a review of the history of this development.

$$\tilde{X}_j^{(n)} := X_j^{(n)} \mathbf{1}_{[|X_j^{(n)}|<\varepsilon_n]}, \qquad j = 1, \ldots, k_n,$$

are also independent and $\sum_{j=1}^{k_n} \tilde{X}_j^{(n)} \to X_t - X_s$ in probability as $n \to \infty$, since the sum is equal to $X_t - X_s$ for all n. The result now follows by an application of the *Lindeberg CLT* to the uniformly bounded variables $\tilde{X}_j^{(n)}$; see (BCPT,[6] p. 78.). ■

Corollary 5.6 *If $\{X_t : t \geq 0\}$ is an additive process with stationary increments having a.s. continuous sample paths, then $\{X_t : t \geq 0\}$ must be Brownian motion.*

The more general Theorem 5.5 may be recast as a characterization of additive processes with continuous sample paths in terms of a deterministic translation and time change of Brownian motion (Exercise 7).

Within this same context of additive processes, the other extreme for sample path structure is that of step functions with positive unit jumps.

Theorem 5.7 *Let $\{X_t : t \geq 0\}$ be a stochastically continuous additive process, almost all of whose sample paths are right-continuous step functions with positive unit jumps. Then, the increments of $\{X_t : t \geq 0\}$ are Poisson distributed.*

Proof Without loss of generality, take $X_0 = 0$ and assume that the paths are not a..s. constant, else regard the distribution of X_t as a degenerate Poisson distribution with mean zero. It is enough to show that for each $t > 0$, X_t has a Poisson distribution with $\mathbb{E}X_t = \lambda(t)$ for some $\lambda(t) > 0$. Partition $(0, t]$ into n intervals of the form $(t_{i-1}, t_i]$, $i = 1, \ldots, n$ having equal lengths $\Delta = t/n$. It is to be noted that explicit dependence of these parameters t_i on the fixed value t, as well as others as they occur, will be suppressed to keep the notation simple. Let $A_i^{(n)} = [X(t_i) - X(t_{i-1}) \geq 1]$ and $S_n = \sum_{i=1}^n \mathbf{1}_{A_i^{(n)}}$ (the number of time intervals with at least one jump occurrence). Let D denote the shortest distance between jumps in the path X_s, $0 < s \leq t$. Then, $P(X_t \neq S_n) \leq P(0 < D \leq t/n) \to 0$ as $n \to \infty$, since $[X_t \neq S_n]$ implies that there is at least one interval containing two or more jumps. Now, by the Poisson approximation (Chapter 1, Lemma 2), we have uniformly over $B \subset \{0, 1, 2, \ldots\}$,

$$|P(S_n \in B) - \sum_{m \in B} \frac{\lambda^m}{m!} e^{-\lambda}| \leq \rho_n \max_{i \leq n} p_i, \tag{5.17}$$

where $p_i \equiv p_i^{(n)} = P(A_i^{(n)}), 1 \leq i \leq n$, $\rho_n = \sum_{i=1}^n p_i$, $\lambda = -\log\theta(t)$, for $\theta(t) := P(X_t = 0) < 1$. Since $\max\{p_i^{(n)} : i = 1, \ldots, n\} \leq \delta_n := P(0 < D \leq \frac{t}{n})$ as $n \to \infty$, we only need to show that

[6] Throughout, BCPT refers to Bhattacharya and Waymire (2016), A Basic Course in Probability Theory.

$$\sum_{i=1}^{n} p_i^{(n)} \to -\log\theta(t) \text{ as } n \to \infty.$$

For this, note that $\theta(t) = \prod_{i=1}^{n} P((A_i^{(n)})^c)$, so that

$$\log\theta(t) = \sum_{i=1}^{n} \log(1 - p_i^{(n)}) = -\sum_{i=1}^{n} p_i^{(n)} - \varepsilon_n,$$

where $0 < \varepsilon_n < \sum_{i=1}^{n} p_i^{(n)2}/(1 - \delta_n) \to 0$ as $n \to \infty$. Hence, $\sum_{i=1}^{n} p_i^{(n)} \to -\log\theta(t)$. Now use Scheffe's theorem for the convergence in total variation for the Poisson distribution with parameter $\sum_{i=1}^{n} p_i^{(n)}$ converging to that parameter λ. ∎

Corollary 5.8 *If $X = \{X_t : t \geq 0\}$ is a stochastically continuous process with stationary independent increments, whose sample paths are right-continuous step functions, then X must be a compound Poisson process. In particular, if the sample paths are right-continuous step functions with positive unit jumps, then X is a (homogeneous) Poisson process.*

Proof Let $N(t) = \#\{s \leq t : X_s - X_{s^-} \neq 0\}$ denote the number of jumps that take place in the time interval 0 to t. For $s < t$, $N(t) - N(s)$ counts the number of jumps in the interval $(s, t]$, which is independent of $N(s)$. In particular, $\{N(t) : t \geq 0\}$ is a process continuous in probability (Exercise 3) with independent increments having sample paths that are right-continuous step functions, which grow by positive unit jumps, with $N(0) = 0$. Thus, $\{N(t) : t \geq 0\}$ must be a (homogeneous) Poisson process by Theorem 5.7. Denoting the successive jump times $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$, we have

$$X_t = X_0 + \sum_{j=1}^{N(t)} \{X_{\tau_j} - X_{\tau_{j-1}}\}.$$

Recall that the inter-arrival sequence $\tau_j - \tau_{j-1}, j \geq 1$ is an i.i.d. sequence of exponentially distributed random variables (see Theorem 4.1). Apply the strong Markov property to see that $\{N(t) : t \geq 0\}$ is independent of the i.i.d. sequence of jump sizes $Y_j = X_{\tau_j} - X_{\tau_{j-i}}, j = 1, 2, \ldots$ (see Corollary 4.2). ∎

Remark 5.4 It follows from methods of Chapter 3 concerning regularization of martingales that the assumed sample path regularity required here is not really a restriction. The regularized process is referred to as the *Lévy modification of X* in the present context.

Throughout $X = \{X_t : t \geq 0\}$ is a given Lévy process defined on a complete[7] probability space $(\Omega, \mathcal{F}, P)$.

Definition 5.3 Let $X = \{X_t : t \geq 0\}$ be a Lévy process defined on a complete probability space $(\Omega, \mathcal{F}, P)$. A collection of σ-fields $\{\mathcal{A}_{st} : 0 \leq s < t, s, t \in [0, \infty)\}$ is called *additive* if

a. $\mathcal{A}_{rs} \vee \mathcal{A}_{st} = \mathcal{A}_{rt}$
b. For $0 \leq t_0 < t_1 < \cdots < t_n$, $\{\mathcal{A}_{t_0,t_1}, \mathcal{A}_{t_1 t_2}, \ldots, \mathcal{A}_{t_{n-1} t_n}\}$ is a collection of independent σ-fields.

For $0 \leq s < t < \infty$, the σ-fields

$$\mathcal{D}_{st}(X) = \sigma\{X_v - X_u : s \leq u < v \leq t\}$$

are referred to as *differential σ-fields* associated with X. We say that X is *adapted* to an additive family $\{\mathcal{A}_{st}\}_{s,t}$ if $X_t - X_s$ is $\mathcal{A}_{st}$-measurable for $s < t$.

Note that if $\{\mathcal{A}_{st}\}$ is additive, then the family of σ-field completions $\{\overline{\mathcal{A}}_{st}\}$ is also additive (Exercise 16). Also, if X is adapted to an additive family $\{\mathcal{A}_{st}\}$, then X is an additive process and $\mathcal{D}_{st}(X) \subset \mathcal{A}_{st}$, provided that $X_0 = 0$ almost surely.

Next, let us analyze the number of jumps of various nonzero sizes which occur in sample paths of a Lévy process X within a specified time t. Since almost all sample paths are right-continuous with left limits, the counting random variables will only be defined up to a subset of $\omega \in \Omega$ of probability zero without further mention of this. The (random) set of jump times is denoted by

$$\Gamma = \Gamma(\omega) := \{t > 0 : X_t(\omega) - X_{t^-}(\omega) \neq 0\}. \tag{5.18}$$

Also we consider the set of jump times and sizes jointly, as given by

$$J = J(\omega) := \{(t, X_t(\omega) - X_{t^-}(\omega)) : t \in \Gamma(\omega)\}. \tag{5.19}$$

Let

$$\mathbb{R}_0 = \mathbb{R} \backslash \{0\}.$$

Note that

$$J(\omega) \subset (0, \infty) \times \mathbb{R}_0$$

is a countable set (Exercise 15).

Let $\mathcal{B}_0$ denote the collection of all Borel subsets of $(0, \infty) \times \mathbb{R}_0$, and define

[7] See BCPT p. 225.

$$\mathcal{R}_0 := \{A \in \mathcal{B}_0 : A \subset (0, \delta) \times \{\Delta : |\Delta| > \delta^{-1}\} \text{ for some } \delta > 0\}.$$

Notice that since δ may be arbitrarily large, membership in $\mathcal{R}_0$ is unrestrictive if one considers fixed positive jump sizes. One may check that $\mathcal{R}_0$ is closed under finite unions and relative complements, i.e., a measure-theoretic *ring* of subsets, but generally not under countable unions nor complements. However, $\mathcal{R}_0$ is also closed under countable intersections (Exercise 18).

Define the possibly infinite valued counting random variables

$$N(A) = N(A, \omega) := \text{card}\{A \cap J(\omega)\}, \quad A \in \mathcal{B}_0. \tag{5.20}$$

Note that for cádlág sample paths, $N(A)$ must be finite for $A \in \mathcal{R}_0$.

Proposition 5.9 *For $A \in \mathcal{B}_0$, either $N(A)$ is a Poisson distributed random variable with finite intensity $\nu(A) := \mathbb{E}N(A)$ or $N(A)$ is $+\infty$ with probability one. Moreover, the random set function $A \to N(A)$, $A \in \mathcal{B}_0$ is a.s. a measure, and the intensity $A \to \nu(A)$ is a measure on $\mathcal{B}_0$. If $A \in \mathcal{R}_0$, then $N(A)$ is Poisson distributed with finite intensity.*

Proof First one must check that $N(A)$ is a random variable; i.e., $\omega \to N(A, \omega)$ is measurable. The proof of measurability is outlined in Exercise 19 where, in fact, one shows that for $A \in \mathcal{R}_0$ and $A \subset (s, t] \times \mathbb{R}_0$, for some $0 \le s < t$, $N(A)$ is measurable with respect to $\overline{\mathcal{D}}_{s,t}(X)$. Consider $A \in \mathcal{R}_0$, and let

$$A_t := A \cap ((0, t] \times \mathbb{R}_0)\,, t \ge 0.$$

The stochastic process $N_t := N(A_t)$ is a right-continuous nondecreasing step function with unit jumps,$N_0 = 0$. Moreover, for any $0 \le s < t$,

$$N_t - N_s = N(A_t \backslash A_s) = N(A \cap ((s, t] \times \mathbb{R}_0))$$

is measurable with respect to the additive differential filtration $\overline{\mathcal{D}}_{s,t}(X)$, and hence is an additive process. Finally, one has continuity in probability at each t, since for any $\varepsilon > 0$, there is a $\delta > 0$ such that

$$P(|N_t - N_s| > \varepsilon) \le P(|X_t - X_s| > \delta^{-1}) \to 0 \text{ as } s \to t.$$

In view of Corollary 5.8, $N_t - N_s$ is Poisson distributed for $0 \le s < t$. Since $A \in \mathcal{R}_0$, it follows that $A = A_t$ for a sufficiently large t, and hence $N(A) = N_t$ has a Poisson distribution with finite intensity. Now, more generally for $A \in \mathcal{B}_0$, there is a nondecreasing sequence $B_1 \subset B_2 \subset \cdots$ in $\mathcal{R}_0$ such that $A = \cup_{n=1}^{\infty} B_n$. Thus, $N(A)$ is the limit of a nondecreasing sequence of Poisson distributed random variables with respective (nondecreasing) intensities $\nu_n = \nu(B_n) < \infty, n \ge 1$. Let $\nu := \lim_n \nu_n$. If $\nu < \infty$, then it follows that $N(A)$ is Poisson distributed with parameter ν. If $\nu = \infty$, then $P(N(A) < \infty) = 0$. Finally, note that for

fixed ω, $A \to N(A, \omega)$ is simply a counting measure, in particular a nonnegative integer valued countably additive set function with $N(\emptyset, \omega) = 0$. It follows from Lebesgue's monotone convergence theorem that $A \to \nu(A) = \mathbb{E}N(A)$ is also a measure on $\mathcal{B}_0$. ■

Remark 5.5 We will make the convention that a Poisson distribution with infinite intensity corresponds to an extended nonnegative integer valued random variable N with $P(N = 0) = 1$.

Definition 5.4 The random measure N will be referred to as the *space-time jump counting measure* associated with X. The intensity measure $\nu(\cdot)$ is referred to as the *Lévy measure* of X.

The random field $N = \{N(A) : A \in \mathcal{B}_0\}$ obtained from the Lévy process X provides counts of the numbers of jumps of specified range of sizes in specified periods of time. From N, one may measure the contributions to the evolution of X arising solely from jumps by introducing the finite sums:

$$M(A) = M(A, \omega) := \sum_{(t,z) \in A \cap J(\omega)} z, \quad A \in \mathcal{R}_0. \tag{5.21}$$

Since

$$M(A) = \lim_{n \to \infty} \sum_k \frac{k}{n} N(A \cap ((0, \infty) \times (\frac{k-1}{n}, \frac{k}{n}])), \tag{5.22}$$

the map $\omega \to M(A, \omega)$ is measurable for each $A \in \mathcal{R}_0$. In fact, it follows that M inherits measurability with respect to the additive differential filtration $\{\overline{\mathcal{D}}_{s,t}(X) : 0 \le s < t < \infty\}$ from that of N; see Exercise 19. A commonly used alternative notation for (5.22) is to write

$$M(A) = \int_{\mathbb{R}_0} \int_{(0,\infty)} z \mathbf{1}_A(t, z) N(dt \times dz), \quad A \in \mathcal{R}_0. \tag{5.23}$$

With this, we may consider the evolution of X apart from jumps by defining

$$Y_t(A) := X_t - M(A_t), \quad A \in \mathcal{R}_0, \quad t \ge 0, \tag{5.24}$$

where as before $A_t = A \cap ((0, t] \times \mathbb{R}_0)$.

Definition 5.5 The measure M in (5.23) will be referred to as the *space-time accumulated jump size measure* associated with X. The processes $\{Y_t(A) : t \ge 0\}$, for $A \in \mathcal{R}_0$, in (5.24) will be referred to as the *jump reduced* processes associated with X.

The following lemma[8] is fundamental to discerning the structure of the Lévy process X.

Lemma 2 (Itô's Fundamental Lemma) *Let $\mathcal{A} = \{\mathcal{A}_{st} : s < t, s, t \in [0, T]\}$ be an additive family of σ-fields, $X = \{X_t\}$ a Lévy process, and $Y = \{Y_t\}$ a Lévy process having Poisson distributed increments, both X and Y being adapted to $\mathcal{A}$. If with probability one there is no common jump point for X and Y, then the two processes are independent. That is, if*

$$P(X_t \neq X_{t^-} \text{ and } Y_t \neq Y_{t^-} \text{ for some } t \in (0, T]) = 0,$$

then $\sigma\{X_t : t \in T\}$ and $\sigma\{Y_t : t \in T\}$ are independent.

Proof It is sufficient to establish independence of $X_t - X_s$ and $Y_t - Y_s$ for fixed $0 \le s < t$. Let $s < t_{n0} < t_{n2} < \cdots < t_{nn} = t$ be a partition of $(s, t]$ into n disjoint subintervals $(t_{n,j-1}, t_{nj}]$, $j = 1, \ldots, n$, of equal length. Define,

$$\begin{aligned} X_{nj} &= X_{t_{nj}} - X_{t_{n,j-1}}, \quad Y_{nj} = Y_{t_{nj}} - Y_{t_{n,j-1}} \\ X_t - X_s &= \sum_{j=1}^{n} X_{nj} \quad Y_t - Y_s = \sum_{j=1}^{n} Y_{nj}. \end{aligned} \tag{5.25}$$

The hypothesis of the lemma implies that as $n \to \infty$, simultaneously for all n, $X_{nj} = 0$ on all intervals where $Y_{nj} > 0$. Thus, there exists $N = N(\omega)$ such that for all $n \ge N(\omega)$, the above is true. Given $\varepsilon > 0$, there then exists N (nonrandom), such that with probability $1 - \varepsilon$, $X_{nj} = 0$ on all intervals, where $Y_{nj} > 0$ for all $n \ge N$. The small probability ε may be thought to be of that of the exceptional j's for which $X_{nj} \neq 0$ but $Y_{nj} > 0$. So, we can assume that these do not occur. We will consider n of the form $n = 2^m$ because as m increases, the exceptional set decreases to the empty set in finite (random) time, under the hypothesis. So for all $n \ge 2^{m_0} = N$ for some positive integer m_0, one has disjoint sets J of intervals, and $\{1, \ldots, n\} \backslash J = I$, say, such that $X_{nj} = 0$, $Y_{nj} > 0$ for $j \in J$ and $Y_{nj} = 0$ for $j \in I$. The sets J and I depend on ω, in general, but one may split the sample space into disjoint sets corresponding to different (nonrandom) sets J of intervals. We may also assume that on those intervals where $Y_{nj} > 0$, it is the case that $Y_{nj} = 1$. This is because the probability that on any one of the n subintervals the event $[Y_{nj} > 1]$ occurs is no more than

$$ne^{-\frac{\lambda(t-s)}{n}} O(\frac{(\lambda(t-s))^2}{n^2}) = O(\frac{1}{n}),$$

[8] See Itô (2008), p. 58, for a more detailed proof of this fundamental lemma.

where λ is the parameter of the Poisson process Y and each of the subinterval lengths is $\frac{t-s}{n}$. Fix an integer $k \geq 0$. We will consider the conditional distribution of $X_t - X_s$ given $Y_t - Y_s = k$. Let n be very large as indicated above. Note that the pairs (X_{nj}, Y_{nj}), $j = 1, \ldots, n$ are i.i.d., in view of the stationary independent increments. The event $Y_t - Y_s = k$ is the disjoint union of $\binom{n}{k}$ sets J corresponding to the choices of k indices j for which $Y_{nj} = 1$, with $Y_{nj} = 0$ for the complementary set I of indices. We consider the conditional distribution of $X_t - X_s$, given such a set J, and show that it does not depend on k or J. Consider the distribution of the two conditionally independent pairs of random variables: $(\sum_{j=1}^n Y_{nj}\mathbf{1}_{[Y_{nj}>0]}, \sum_{j=1}^n X_{nj}\mathbf{1}_{[Y_{nj}>0]})$, $(\sum_{j=1}^n Y_{nj}\mathbf{1}_{[Y_{nj}=0]}, \sum_{j=1}^n X_{nj}\mathbf{1}_{[Y_{nj}=0]})$. The event $[Y_t - Y_s = k]$ means that the first pair is $(k, \sum_{j\in J} X_{nj}\mathbf{1}_{[Y_{nj}>0]})$ and the second pair is $(0, \sum_{j=1}^n X_{nj}\mathbf{1}_{[Y_{nj}=0]})$, where J is the index set of k intervals, in which $Y_{nj} = 1$, and I is the complementary set $\{1, 2, \ldots, \}\backslash J$. On the set J, $\sum_{j\in J} X_{nj} = o(1)$ in probability, denoted $o_p(1)$, because it is the sum of k i.i.d random variables, each of the order $o_p(1/n)$. This last fact makes use of the hypothesis of the lemma, which implies (see Remark 5.6)

$$P(X_{nj} \neq 0, Y_{nj} > 0) = o_p(1/n).$$

Hence, conditionally given $Y_t - Y_s = k$, $\sum_{j\in I} X_{nj}\mathbf{1}_{[Y_{nj}=0]}$ is, asymptotically in probability, the same as the sum $\sum_{j=1}^n X_{nj} = X_t - X_s$. In terms of conditional expectations, this may be expressed as

$$\mathbb{E}(\exp\{i\xi \sum_{j=1}^{n} X_{nj}\}|J) - \mathbb{E}(\exp\{i\xi \sum_{j\in I} X_{nj}\}|J) \to 0 \tag{5.26}$$

in probability as $n \to \infty$ and $\mathbb{E}(\exp\{i\xi \sum_{j\in I} X_{nj}\}|J) = \mathbb{E}\exp\{i\xi \sum_{j\in I} X_{nj}\}$. This shows that the conditional distribution of $X_t - X_s = \sum_{j=1}^n X_{nj}$ is asymptotically the same no matter what $k = Y_t - Y_s$ is. This implies $X_t - X_s$ is independent of $Y_t - Y_s$. ■

Remark 5.6 Let $\Theta_n = P(X_{nj}\mathbf{1}_{[Y_{nj}>0]} \neq 0)$. Then,

$$P(X_{nj} = 0, Y_{nj} > 0) \leq 1 - \Theta_n,$$

so that

$$P(X_{nj} = 0 \text{ for for all } j \text{ such that } Y_{nj} > 0) \leq (1 - \Theta_n)^n,$$

which converges, as $n \to \infty$, to

$$P(\text{In the interval } [s, t] \text{ jumps of } X \text{ and } Y \text{ take place on disjoint sets}) = 1,$$

by hypothesis, unless $X_t - X_s = 0$ with probability one (a case in which the independence is trivially true). This implies $\Theta_n = o(1/n)$. One may perhaps also argue more simply that the event $P(X_{nj} \neq 0) = 1 - \exp\{-\gamma(t-s)/n\} = O(1/n)$, where γ is the parameter of the exponential jump distribution of X. But the probability that X and Y have both jumps in the interval $[(j-1)(t-s)/n, j(t-s)/n]$ is of smaller order than $O(1/n)$.

Let

$$A_t := A \cap ((0,t] \times \mathbb{R}_0)\,, t \geq 0.$$

The stochastic process

$$N_t \equiv N_t(A) := N(A_t) \tag{5.27}$$

is a.s. a right-continuous nondecreasing step function with unit jumps, $N_0 = 0$.

Proposition 5.10 *For each $A \in \mathcal{R}_0$, the jump reduced process $\{Y_t(A) : t \geq 0\}$ associated with X by (5.24) is a Lévy process and is independent of the Poisson process $\{N_t(A) : t \geq 0\}$.*

Proof Observe that for fixed $A \in \mathcal{R}_0$, $\{Y_t(A) : t \geq 0\}$ is a Lévy process and $\{N_t(A) : t \geq 0\}$ is a Poisson process with step function sample paths having unit jumps. Moreover, these two processes are adapted to a common additive differential filtration $\{\overline{\mathcal{D}}_{s,t} : 0 \leq s < t < \infty\}$. Clearly, by sample path, the two processes cannot have a common Jump time. Thus, the assertion follows from Lemma 2. ■

Proposition 5.11 *For disjoint $A^1, \ldots, A^n \in \mathcal{R}_0, n \geq 1$, the stochastic processes $N^i = \{N_t^i \equiv N(A_t^i) : t \geq 0\}$, $i = 1, 2, \ldots n$ and $\{Y_t(\cup_{j=1}^n A^j) : t \geq 0\}$ are mutually independent. Moreover, this independence statement is also valid for the corresponding space-time accumulated jump size processes $M^i, i = 1, 2, \ldots n$, in place of the jump count processes $N^i, i = 1, 2, \ldots$.*

Proof A simple recursive relationship provides the key to tracking independence. Observe that if one denotes the jump reduced (hence Lévy) process $Y(A) = \{Y_t(A) : t \geq 0\}$ associated with $X = \{X_t : t \geq 0\}$ by $Y_X(A)$ and similarly denote the space-time jump counting measure associated with X by N_X, then, for the processes $N^i, Y_t^i := Y_t(\cup_{j=1}^i A^j), i = 1, 2 \ldots n$, one has a recursive relationship given by

$$N_t^{i+1} = N_{Y^i}(A_t^{i+1}), \quad \text{and} \quad Y_t^{i+1} = (Y_{Y^i})_t(A^{i+1}),\ i = 0, 1 \ldots, n-1,$$

where $Y^0 = X$. Thus, one has, denoting the *join*[9] of σ-fields by $\vee$, $\sigma(N^{i+1}) \vee \sigma(Y^{i+1}) \subset \sigma(Y^i), i = 0, 1, \ldots, n-1$, and hence

[9] The join refers to the σ-field generated by a union of σ-fields.

$$\sigma(N^{i+1}) \vee \sigma(N^{i+1}) \vee \cdots \vee \sigma(N^n) \vee \sigma(Y^n) \subset \sigma(Y^i), \quad i = 0, 1, \ldots, n-1.$$

Applying the previous proposition (i.e., for $n = 1$) to the Lévy process $\{Y_t^i(A^{i+1}) : t \geq 0\}$, it follows that $\sigma(N^{i+1})$ is independent of $\sigma(Y^{i+1})$. Now let $F_i \in \sigma(N^i)$, $i = 1, 2, \ldots, n$ and $G \in \sigma(Y^n)$. Then $F_{i+1} \cap \cdots \cap F_n \cap G \in \sigma(Y^i)$ for each $i = 0, 1 \ldots, n-1$. Therefore,

$$P(\cap_{j=1}^n F_j \cap G) = P(F_1)P(\cap_{j=2}^n F_j \cap G) = \cdots = \prod_{j=1}^n P(F_j)P(G).$$

That is, $\sigma(N^1), \ldots, \sigma(N^n), \sigma(Y^n)$ are independent. The independence readily extends to the corresponding accumulated jump size processes M^i, $i = 1, 2, \ldots n$, in place of the jump count processes N^i, $i = 1, 2, \ldots$, via the a.s. limit formulae

$$M^i(A) = \lim_{n\to\infty} \sum_k \frac{k}{n} N^i(A \cap (0, \infty) \times (\frac{k-1}{n}, \frac{k}{n}]), i = 1, \ldots, n.$$

■

Corollary 5.12 *For any disjoint $A^1, \ldots, A^n$ in $\mathcal{B}_0$, $n \geq 1$, the Poisson random variables $N(A^1), \ldots, N(A^n)$ are independent.*

Proof First, consider the case that each $A^i \in \mathcal{R}_0$, $i = 1, 2, \ldots n$. Then, for sufficiently large t one has $A^i = A_t^i$, $i = 1, \ldots, n$. So the assertion is immediate in this case. More generally, for any $A^i \in \mathcal{B}_0$ there is a nondecreasing sequence $B_{i,1} \subset B_{i,2} \subset \cdots$ in $\mathcal{R}_0$ such that $A^i = \cup_{n=1}^\infty B_{i,n}$. Thus, each $A^i \in \mathcal{A}_i := \sigma(B_{i,1}, B_{i,2}, \ldots)$, $i = 1, 2, \ldots n$ and $\mathcal{A}_i$, $i = 1, 2, \ldots, n$ are independent σ-fields. ■

Proposition 5.13 *For $A \in \mathcal{R}_0$,*

$$\mathbb{E}e^{i\xi M(A)} = \exp\{-\int\int_A (1 - e^{i\xi z})\nu(dt \times dz)\}, \quad \xi \in \mathbb{R}.$$

If furthermore $A \subset \{(t, z) : |z| < a\}$ for some $0 < a < \infty$, then

$$\mathbb{E}M(A) = \int\int_A z\nu(dt \times dz), \quad \textit{and} \quad \mathrm{Var}(M(A)) = \int\int_A z^2\nu(dt \times dz).$$

Proof For $A \in \mathcal{R}_0$ and integer $n \geq 1$, let

$$A_k^n := \{(t, z) \in A : \frac{k}{n}a < z \leq \frac{k+1}{n}a\},$$

so that a.s.

$$M(A) = \int\int_A zN(ds \times dz) = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{k}{n} N(A_k^n).$$

Now, $A_i^n \cap A_j^n = \emptyset$ for $i \neq j$ and hence

$$\begin{aligned}
\mathbb{E}e^{i\xi M(A)} &= \lim_{n\to\infty} \mathbb{E}e^{i\xi \sum_k a \frac{k}{n} N(A_k^n)} \\
&= \lim_{n\to\infty} \prod_k \exp\{\nu(A_k^n)(e^{i\xi \frac{k}{n}} - 1)\} \\
&= \lim_{n\to\infty} \exp\{\sum_k (e^{i\xi \frac{k}{n}} - 1)\nu(A_k^n)\} \\
&= \exp\{\int\int_A (e^{i\xi z} - 1)\nu(dt \times dz)\}. \qquad (5.28)
\end{aligned}$$

The asserted formulae for the mean and variance follow from the Taylor expansion of the exponential appearing in the integrand of the formula for the cumulant expansion of $\log \mathbb{E}e^{i\xi M(A)}$. ■

Observe that from continuity in probability, it follows that there can be no fixed time t of discontinuity, i.e.,

$$N(\{t\} \times \mathbb{R}_0) = 0 \text{ a.s. for each t.}$$

Taking expectations, therefore, one has

$$\nu(\{t\} \times \mathbb{R}_0) = 0.$$

Also since for any $\delta > 0$, $A = (0, t] \times \{z : |z| > \delta\} \in \mathcal{R}_0$. It follows from Proposition 5.9 that

$$\mathbb{E}N(A) = \int\int_A \nu(dt \times dz) < \infty.$$

On the other hand, for "small jumps," one also has the following.

Proposition 5.14 *For* $A = \{(t, z) : 0 \le t \le s, |z| < 1\}$,

$$\int\int_A z^2 \nu(dt \times dz) < \infty.$$

Proof Let $t, \delta > 0$ and consider $J_s(\delta) = \{(s, z) : 0 < s \le t, \delta < |z| < 1\} \in \mathcal{R}_0$. Since $\{M(J_s(\delta)) : 0 < s \le t\}$ and $\{X_s - M(J_s(\delta)) : 0 < s \le t\}$ are independent and $J_t(\delta) = J(\delta)$, it follows that $M(J(\delta))$ and $X_t - M(J(\delta))$ are independent. Thus,

$$\mathbb{E}e^{i\xi X_t} = \mathbb{E}e^{i\xi M(J(\delta))}\mathbb{E}e^{i\xi\{X_t - M(J(\delta))\}}.$$

Thus, using Proposition 5.13 and the bound $\cos(x) \le 1 - \frac{x^2}{4}$, for $|x| \le 1$,

$$\begin{aligned}|\mathbb{E}e^{i\xi X_t}| \le |\mathbb{E}e^{i\xi M(J(\delta))}| &= \exp\{\int\int_{J(\delta)} (\cos(\xi z) - 1)\nu(ds \times dz)\} \\ &\le \exp\{-\frac{\xi^2}{4}\int\int_{J(\delta)} z^2\nu(ds \times dz)\}, \quad |\xi| < 1.\end{aligned}$$

If

$$\int\int_{\{(s,z):0<s\le t, 0<|z|<1\}} z^2\nu(ds \times dz) = \infty,$$

then $|\mathbb{E}e^{i\xi X_t}| \le 0$ for $0 < |\xi| < 1$. But letting $\xi \to 0$, one sees that this is impossible. Thus, the integral must be finite as asserted. ■

The jump process M can be used to enumerate the contribution to the evolution of X due purely to jumps of magnitudes between $\frac{1}{k+1}$ and $\frac{1}{k}$ for $k = 0, 1, \ldots$ ($\frac{1}{0} = \infty$). It is also straightforward to consider positive and negative jumps separately. First, let us consider positive jumps and successively define

$$\begin{aligned}M_{k+1}(t) &:= M((0, t] \times [\frac{1}{k+1}, \frac{1}{k})) \\ &= \int\int_{(0,t]\times[\frac{1}{k+1},\frac{1}{k})} zN(ds \times dz), \quad k = 0, 1, \ldots, t \ge 0.\end{aligned}$$

That is, $M_{k+1}(t)$ sums up the jump increases $X_s - X_{s^-}$ of size between $\frac{1}{k+1}$ and $\frac{1}{k}$ at time points $0 < s \le t$, where such jumps occur. Define

$$S_k(t) := M_1(t) + \sum_{j=2}^{k}\{M_j(t) - \mathbb{E}M_j(t)\}, \quad t \ge 0, k = 1, 2, \ldots.$$

Proposition 5.15 *$\{M_n(t) : t \ge 0\}$ are mutually independent Lévy processes for $n = 1, 2, \ldots$, each having sample paths, which are step functions. Also, for each $n \ge 1$, $S_n := \{S_n(t) : t \ge 0\}$ is a Lévy process with alternate representation of the form*

$$S_n(t) = \int\int_{(0,t]\times[1/n,\infty)} zN(ds \times dz) - \int\int_{(0,t]\times[1/n,1)} z\nu(ds \times dz), t \ge 0.$$

Proof The first assertion is a direct consequence of what has already been proven. Since for each $n \geq 0$,

$$\mathbb{E}M_{n+1}(t) = \int\int_{\{(s,z):0<s\leq t, \frac{1}{n}\leq z<\frac{1}{n+1}\}} z\nu(ds \times dz)$$

is continuous in t, it follows that each $\{S_n(t) : t \geq 0\}$ is also a Lévy process for $n \geq 1$ with

$$\begin{aligned} &S_n(t) \\ &= M((0,t] \times [1,\infty)) + \sum_{j=2}^{n}\{M((0,t] \times [\frac{1}{j}, \frac{1}{j-1})) \\ &\quad - \int\int_{\{(s,z):0<s\leq t, \frac{1}{j}\leq z<\frac{1}{j-1}\}} z\nu(ds \times dz)\} \\ &= M((0,t] \times [\frac{1}{n}, \infty)) - \int\int_{\{(s,z):0<s\leq t, \frac{1}{n}\leq |z|<1\}} z\nu(ds \times dz). \end{aligned}$$

The alternate representation now follows from the integral formula (5.23) for M. ■

Using appropriate maximal inequalities, we will see that with probability one the processes S_n converge uniformly on compact time intervals to a process S_∞. In particular, therefore, S_∞ will also a.s. have cádlág paths. Unfortunately, the Lèvy measure only appears in connection with the increases in X at jumps of all sizes between 0 and 1. In particular, the Lèvy measure for any jumps larger than one cannot be included, because the natural centering term $\int_{(0,t]\times[1,\infty)} z\nu(ds \times dz)$ for S_1 may diverge. This may be remedied by a re-centering "trick" along the following lines: Recall that

$$\int\int_{(0,t]\times[1,\infty)} \nu(ds \times dz) < \infty \quad \text{and} \quad \int\int_{(0,t]\times(0,1)} z^2\nu(ds \times dz) < \infty. \tag{5.29}$$

The first of these is because $(0,t] \times [1,\infty) \in \mathcal{R}_0$ and $N(A) < \infty$ a.s. for $A \in \mathcal{R}_0$, and the second was established in Proposition 5.14. Since $\frac{z^2}{1+z^2} \sim 1$ as $|z| \to \infty$ and $\frac{z^2}{1+z^2} \sim z^2$ for $z \to 0$, one may more simply write

$$\int\int_{\{(s,z):0<s\leq t, z\neq 0\}} \frac{z^2}{1+z^2}\nu(ds \times dz) < \infty.$$

The trick is to replace the centering term $\int\int_{(0,t]\times[1/n,1)} z\nu(ds \times dz)$ in the formula for S_n by $\int\int_{(0,t]\times[1/n,\infty)} \frac{z}{1+z^2}\nu(ds \times dz)$, since $\frac{z}{1+z^2} \sim z$ as $z \to 0$, i.e., define

$$X_n^+(t) := \int\int_{(0,t]\times[1/n,\infty)} zN(ds\times dz) - \int\int_{(0,t]\times[1/n,\infty)} \frac{z}{1+z^2}\nu(ds\times dz). \tag{5.30}$$

Then, since $z - \frac{z}{1+z^2} = \frac{z^3}{1+z^2}$, the difference is

$$X_n^+(t) - S_n(t) = \int\int_{(0,t]\times[1/n,1)} \frac{z^3}{1+z^2}\nu(ds\times dz) - \int\int_{(0,t]\times[1,\infty)} \frac{z}{1+z^2}\nu(ds\times dz).$$

Now, in view of (5.29) and since $|z|^3 < z^2$ for $|z| < 1$ and $|z| < z^2$ for $z > 1$, this difference converges uniformly on compact time intervals as $n\to\infty$. While $X_n^+, n\geq 1$, does not have the natural centering from the point of view of using maximal inequalities to show a.s. uniform convergence on compacts, it *does* explicitly contain the Lèvy measure corresponding to (positive) jumps of all sizes. However, since $S_n, n\geq 1$, is naturally centered (from the point of view of maximal inequalities), we can show that S_n a.s. converges uniformly on compact time intervals, and hence X_n^+ will so converge as well ! It suffices to show that the sequence $\{S_n : n\geq 1\}$ is a.s. uniformly Cauchy on compact time intervals $[0,T]$. That is, in view of the definition of S_n, we wish to show a.s. as $m,n\to\infty$,

$$\sup_{0\leq t\leq T} |\sum_{k=m+1}^{n}\{M_k(t) - \mathbb{E}M_k(t)\}| \to 0. \tag{5.31}$$

The proof will be obtained by the following version of a Kolmogorov's maximal inequality for continuous parameter processes.

Lemma 3 *For arbitrary $m < n$, $T > 0$ and $\varepsilon > 0$,*

$$P(\sup_{0\leq t\leq T} |\sum_{k=m+1}^{n}\{M_k(t) - \mathbb{E}M_k(t)\}| > \varepsilon) \leq \frac{\mathbb{E}|\sum_{k=m+1}^{n}\{M_k(T) - \mathbb{E}M_k(T)\}|^2}{\varepsilon^2}.$$

Proof Since $t\to\sum_{k=m+1}^{n}\{M_k(t) - \mathbb{E}M_k(t)\}$ is a Lévy process,

$$\begin{aligned} &P(\sup_{0\leq t\leq T} |\sum_{k=m+1}^{n}\{M_k(t) - \mathbb{E}M_k(t)\}| > \varepsilon) \\ &= \lim_{N\to\infty} P(\max_{0\leq j\leq N} |\sum_{k=m+1}^{n}\{M_k(\frac{jT}{N}) - \mathbb{E}M_k(\frac{jT}{N})\}| > \varepsilon). \end{aligned} \tag{5.32}$$

Let $X_j = M_k(\frac{jT}{N}) - \mathbb{E}M_k(\frac{jT}{N}) - M_k(\frac{(j-1)T}{N}) + \mathbb{E}M_k(\frac{(j-1)T}{N})$, $j = 1,\dots N$. Then, $X_1,\dots,X_N$ are independent with mean zero, and thus, by Kolmogorov's maximal inequality,

$$P(\max_{j\le N}|\sum_{i=1}^{j}X_i|>\varepsilon)\le\frac{1}{\varepsilon^2}\mathbb{E}|\sum_{j=1}^{N}X_j|^2.$$

Inserting this into the limit establishes the asserted inequality. ■

Note that the second moment appearing in the upper bound may be expressed in terms of the variance formula for M obtained in Proposition 5.13 as

$$\begin{aligned}\beta(m,n,T,\varepsilon) &:= \frac{1}{\varepsilon^2}\mathbb{E}|\sum_{k=m+1}^{n}\{M_k(T)-\mathbb{E}M_k(T)\}|^2\\ &=\frac{1}{\varepsilon^2}\int\int_{\{(s,z):0\le s\le T,\frac{1}{n}\le z\le\frac{1}{m}\}}z^2\nu(ds\times dz).\end{aligned}\tag{5.33}$$

Since $\int\int_{\{(s,z):0<s\le T,|z|\le 1\}}z^2\nu(ds\times dz)<\infty$, one has that

$$\lim_{m,n\to\infty}\beta(m,n,T,\varepsilon)=0.$$

This gives uniformly Cauchy partial sums on compact time intervals in probability. The stronger statement that with probability one S_n converges uniformly on every compact time interval will be achieved with the following refinement of the maximal inequality.

Lemma 4 *For $\beta(m,n,T,\varepsilon)<1/2$, one has*

a. $P(\max_{m<j\le n}\sup_{0\le t\le T}|\sum_{k=m+1}^{j}\{M_k(t)-\mathbb{E}M_k(t)\}|>2\varepsilon)\le 2\beta(m,n,T,\varepsilon)$
b. $P(\max_{m<i<j\le n}\sup_{0\le t\le T}|\sum_{k=i+1}^{j}\{M_k(t)-\mathbb{E}M_k(t)\}|>4\varepsilon)\le 4\beta(m,n,T,\varepsilon)$

Proof The second inequality is implied by the first. Let

$$X_{ij}:=\sup_{0\le t\le T}|\sum_{k=i+1}^{j}\{M_k(t)-\mathbb{E}M_k(t)\}|$$

and note that

$$X_{ik}\ge X_{ij}-X_{jk},\quad \text{for } i<j<k.\tag{5.34}$$

For fixed m, let

$$A_j{:=}[X_{m,m+1}\le 2\varepsilon,\ldots,X_{m,j-1}\le 2\varepsilon,X_{m,j}>2\varepsilon],\ \ B_j{=}[X_{j,n}\le\varepsilon],\ j=m{+}1,\ldots,n,$$

where $B_n:=\Omega$. Then, by the first lemma,

$$P(B_j)=1-P(Y_{jn}>\varepsilon)\ge 1-\beta(j,n,T,\varepsilon)\ge 1-\beta(m,n,T,\varepsilon)>1/2,$$

and using (5.34), $\cup_{j=m+1}^{n} A_j \cap B_j \subset [X_{m,n} > \varepsilon]$, so that, since the events $A_j \cap B_j$, $m+1 \leq j \leq n$, are disjoint, one has by the previous lemma that

$$\sum_{j=m+1}^{n} P(A_j \cap B_j) \leq \beta(m,n,T,\varepsilon).$$

But the events A_j and B_j are independent and $P(B_j) > 1/2$. Thus,

$$\begin{aligned}\frac{1}{2}P(\cup_{j=m+1}^{n} A_j) &\leq \frac{1}{2}\sum_{j=m+1}^{n} P(A_j)\\ &\leq \sum_{j=m+1}^{n} P(A_j)P(B_j) = \sum_{j=m+1}^{n} P(A_j \cap B_j)\\ &\leq \beta(m,n,T,\varepsilon). \end{aligned} \tag{5.35}$$

Since $[\max_{m\leq j\leq n} X_{mj} > 2\varepsilon] \subset \cup_{j=m+1}^{n} A_j$ the proof of the lemma is complete. ■

Although $\sup_{0<t\leq T} |\sum_{k=m+1}^{n}\{M_k(t) - \mathbb{E}M_k(t)\}|$ is neither decreasing nor increasing with m or n, respectively, one can achieve monotonic dependence by considering $\max_{m<i<j\leq n}\sup_{0<t\leq T} |\sum_{k=i+1}^{j}\{M_k(t) - \mathbb{E}M_k(t)\}|$. In particular,

$$\sup_{0<t\leq T} |\sum_{k=m+1}^{n}\{M_k(t) - \mathbb{E}M_k(t)\}| \leq \max_{m<i<j\leq n}\sup_{0<t\leq T} |\sum_{k=i+1}^{j}\{M_k(t) - \mathbb{E}M_k(t)\}|$$

and $\lim_{m,n\to\infty}\max_{m<i<j\leq n}\sup_{0<t\leq T} |\sum_{k=i+1}^{j}\{M_k(t) - \mathbb{E}M_k(t)\}|$ exists (but may be infinite).

Proposition 5.16 *With probability one, the processes $S_n = \{S_n(t) : t \geq 0\}$ and $X_n^+ = \{X_n^+(t) : t \geq 0\}$ converge uniformly on compact time intervals to a Lévy process X^+.*

Proof The convergence for X_n^+ follows from that of S_n and the difference $X_n^+ - S_n$. The sample path structure follows from the uniformity of convergence on compacts. So it suffices to prove the convergence for S_n. Let $\varepsilon > 0$ and apply the second inequality of the previous lemma to write

$$\begin{aligned}&P(\lim_{m,n\to\infty}\max_{m<i<j\leq n}\sup_{0<t\leq T} |\sum_{k=i+1}^{j}\{M_k(t) - \mathbb{E}M_k(t)\}| > 4\varepsilon)\\ &\leq \lim_{m,n\to\infty} P(\max_{m<i<j\leq n}\sup_{0<t\leq T} |\sum_{k=i+1}^{j}\{M_k(t) - \mathbb{E}M_k(t)\}| > 4\varepsilon)\end{aligned}$$

$$\leq \lim_{m,n\to\infty} 4\beta(m,n,T,\varepsilon) = 0.$$

Thus, the a.s. uniformity is achieved. ■

The negative jumps may be handled similarly by defining

$$X_n^-(t) := \int\int_{(0,t]\times(-\infty,-1/n]} zN(ds\times dz) - \int\int_{(0,t]\times(-\infty,-1/n]} \frac{z}{1+z^2}\nu(ds\times dz). \tag{5.36}$$

The idea now is that by removing jumps of all sizes, i.e., a.s. defining

$$Z_t = X_t - X_t^+ - X_t^-, \quad t \geq 0,$$

one is left with a Lévy process $Z = \{Z_t : t \geq 0\}$ having a.s. continuous sample paths and, therefore, Gaussian increments.

Theorem 5.17 (Lévy-Itô Decomposition) *Let $X = \{X_t : t \geq 0\}$ be a Lévy process. Then there is (i) a Poisson random measure $N = \{N(A) : A \in \mathcal{B}_0\}$ with intensity $\nu(A) = \mathbb{E}N(A)$ such that $t \to \int\int_{(0,t]\times\mathbb{R}_0} \frac{z^2}{1+z^2}\nu(ds\times dz)$ is finite and continuous, and (ii) a Lèvy process $Z = \{Z_t : t \geq 0\}$ with a.s. continuous paths and Gaussian increments having mean $m(t) = \mathbb{E}Z_t$ and variance $v(t) = \mathrm{Var}(Z_t)$ and Z is independent of the Poisson random measure N, such that a.s.*

$$X_t = Z_t + \lim_{n\to\infty}\{\int\int_{R_n(t)} zN(ds\times dz) - \int\int_{R_n(t)} \frac{z}{1+z^2}\nu(ds\times dz)\},$$

where $R_n(t) := \{(s,z) : 0 < s \leq t, |z| > 1/n\}$.

Proof The process Z is adapted to the additive differential filtration of X and $Z_0 = 0$, and therefore, Z is an additive process. Also, with probability one

$$Z_t = \lim_{n\to\infty}(X_t - X_n^+(t) - X_n^-(t))$$

a.s. uniformly on compact intervals. Moreover,

$$\begin{aligned} &X_n^+(t) + X_n^-(t) \\ &= \int\int_{\{(s,z):0<s\leq t,|z|>\frac{1}{n}\}} zN(ds\times dz) - \int\int_{\{(s,z):0<s\leq t,|z|>\frac{1}{n}\}} \frac{z}{1+z^2}\nu(ds\times dz). \end{aligned}$$

The second integral term is continuous in t. For every $n > m$, the processes $Z_n(t) := X_t - X_n^+(t) - X_n^-(t)$ a.s. have no jumps of absolute size larger than $\frac{1}{m}$. Thus, it follows that the limiting process Z a.s. has continuous paths. The asserted independence of the two processes can be seen as follows. $Z_n(t)$ differs from $X_t - M((0,t]\times\{z : |z| \geq \frac{1}{n}\})$ by a deterministic continuous function of t

and hence is independent of $\{N(A) : A \in \mathcal{R}_0, A \subset \{(t,z) : |z| \geq \frac{1}{m}\}$ for $n > m$. Thus, passing to the a.s. limit, the independence follows by standard measurability arguments. ■

A converse may be obtained for the following three suitably prescribed components: mean $m(t) = \mathbb{E}Z_t$, variance $v(t) = \mathrm{Var}(Z_t)$, and Lévy measure $\nu(dt \times dz)$ as follows.

Theorem 5.18 (Converse Lévy-Itô Decomposition) *Suppose that (a): m is a continuous function on $[0,\infty)$ with $m(0) = 0$, (b): v is a continuous nondecreasing function on $[0,\infty)$ with $v(0) = 0$, and (c): $\nu(dt \times dz)$ is a measure on $(0,\infty) \times \mathbb{R}\backslash\{0\}$ such that $\int\int_{\{(s,z):0<s\leq t,|z|>0\}} \frac{z^2}{1+z^2}\nu(ds \times dz) < \infty$ and $\nu(\{t\} \times \mathbb{R}\backslash\{0\}) = 0$ for each $t > 0$. Then, there is an a.s. unique Lèvy process whose Lévy-Itô decomposition has these three components.*

Proof We outline the construction of the Lévy process X and leave the details as exercises. The construction can be accomplished by constructing the Gaussian process and Poisson random measure on two different probability spaces and then take the product space for their independent joint specification. More specifically, for Z let

$$Z_t = m(t) + B(\sqrt{v(t)}), t \geq 0,$$

where $\{B(s) : s \geq 0\}$ is standard Brownian motion starting at 0. For the construction of the Poisson random measure N with intensity ν, write $(0,\infty) \times \mathbb{R}\backslash\{0\} = \cup_{n=1}^{\infty} D_n$, where $D_1, D_2, \ldots$ are disjoint sets in $\mathcal{R}_0$ having $0 < \nu(D_n) < \infty$. Normalize ν restricted to D_n to a probability $\mu_n(A) = \nu(A)/\nu(D_n)$ for $A \in D_n \cap \mathcal{B}_0$. Let Y_n be a Poisson distributed random variable with intensity $\nu(D_n)$ and independent of an i.i.d. sequence $X_1^n, X_2^n, \ldots$ in D_n with distribution μ_n, and independently for $n = 1, 2, \ldots$. Define

$$N(A) = \sum_{n=1}^{Y_n} \mathrm{card}\{j : X_j^n \in A\}, \quad A \in \mathcal{B}_0.$$

From here, one may directly use the Lévy-Itô decomposition formula to define the process X, and then directly check that it is a Lévy process. To prove uniqueness, simply note that for each t the Gaussian random variable Z_t is uniquely determined by its mean $m(t)$ and variance $v(t)$, respectively. Now by the Lévy-Itô decomposition in terms of the a.s. limit,

$$X_t - Z_t = \lim_{n\to\infty}\{\int\int_{R_n(t)} zN(ds \times dz) - \int\int_{R_n(t)} \frac{z}{1+z^2}\nu(ds \times dz)\}.$$

Recalling the characteristic function of $M_n(t) - M_n(s) = \int\int_{R_n(s,t)} zN(d\tau \times dz)$ as

$$\mathbb{E}e^{i\xi(M_n(t)-M_n(s))} = \exp\{\int\int_{R_n(s,t)} (e^{i\xi z} - 1)\nu(d\tau \times dz)\},$$

it follows that the distribution of the Lévy process $\{M_n(t) : t \geq 0\}$ is uniquely determined by $\nu(dt \times dz)$. ■

As noted in Chapter 2, Example 7, stationary, independent increments make the homogeneous Lévy process a very special Markov process, whose transition probabilities may be expressed as

$$p(t; x, B) = Q_t(B - x), \quad t \geq 0,\ B \in \mathcal{B},\ x \in \mathbb{R} \tag{5.37}$$

where $\{Q_t(dy) : t \geq 0\}$ is a *convolution semigroup* of probabilities on $(\mathbb{R}, \mathcal{B})$ in the sense that

$$Q_{t+s} = Q_t * Q_s, \qquad s, t \geq 0. \tag{5.38}$$

Recall

Definition 5.6 A probability measure Q on $(\mathbb{R}^k, \mathcal{B}^k)$ is said to be infinitely divisible if, for each integer $n \geq 1$, there is a probability measure Q_n such that $Q = Q_n^{*n}$.

As a result of (5.38), the distribution Q_t of the increment $X_{t+s} - X_s$ must be *infinitely divisible*. It is less obvious that this extends to arbitrary Lévy processes, homogeneous or not.

Corollary 5.19 (Lévy-Khinchine Formula) *Let $X = \{X_t : t \geq 0\}$ be a Lévy process with the indicated Lévy-Itô decomposition with*

$$m(t) := \mathbb{E}Z_t, \quad v(t) := Var Z_t \quad \nu(A) := \mathbb{E}N(A), \quad A \in \mathcal{B}_0.$$

Then, for any $0 < s < t$, the increment $X_t - X_s$ has an infinitely divisible distribution with characteristic function given by

$$\begin{aligned}&\mathbb{E}e^{i\xi(X_t-X_s)}\\ &= \exp\{i\xi(m(t) - m(s)) - \frac{v(t) - v(s)}{2}\xi^2 + \int\int_{(s,t]\times\mathbb{R}_0} (e^{i\xi z} - 1 - \frac{i\xi z}{1+z^2})\nu(d\tau \times dz)\}.\end{aligned}$$

Proof For $0 < s < t$ write

$$X_t - X_s = Z_t - Z_s + \lim_{n\to\infty}\{\int\int_{R_n(s,t)} zN(d\tau \times dz) - \int\int_{R_n(s,t)} \frac{z}{1+z^2}\nu(d\tau \times dz)\},$$

where $R_n(s, t) := \{(\tau, z) : s < \tau \leq t, |z| \geq \frac{1}{n}\}$, and $Z = \{Z_t : t \geq 0\}$ is a Lévy process with continuous paths and Gaussian increments independent of the Poisson rondom measure N. From the independence of increments, one has for $s \leq t$,

$$v(t) = Var(Z_t - Z_s + Z_s) = Var(Z_t - Z_s) + v(s).$$

The formula follows using standard properties of characteristic functions for independent sums, a.s. limits (via Lebesgue's dominated convergence theorem), and formulae for Gaussian and compound Poisson characteristic functions. The infinite divisibility is obtained by the converse to the Lévy-Itô decomposition upon noting that any r-th root ($r = 1, 2, \dots$) of the characteristic function is of the same form with m and V replaced by $m_r = m/r$, $V_r = V/r$ and n replaced by n/r. ■

Thus, the structure of the class of infinitely divisible distributions is fully embodied in the distribution of Lévy processes, and conversely. Since every additive process, which is continuous in probability, admits a martingale regularization, which makes it a Lévy process, the Lévy-Khinchine formula extends accordingly to the increments of such processes.

The homogeneous Lévy processes, i.e., also having stationary increments, enjoy further special structure related to the special class of infinitely divisible laws, the *stable laws* being an important further special case (see Examples 2, 3).

Theorem 5.20 *A Lévy process $X = \{X_t : t \geq 0\}$ is homogeneous if and only if*

$$m(t) = mt, \quad v(t) = vt, \quad \nu(dt \times dz) = dt \times \tilde{\nu}(dz),$$

for some constants $m \in \mathbb{R}$, $v \geq 0$, and measure $\tilde{\nu}$ on $\mathbb{R}_0$ such that

$$\int_{\mathbb{R}_0} \frac{z^2}{1+z^2} \tilde{\nu}(dz) < \infty.$$

Proof Since

$$X_t = Z_t + \lim_{n\to\infty} M_n(t)$$

where each $\{M_n(t) : t \geq 0\}$ is clearly homogeneous Lévy process, it follows that Z must also be homogeneous. In particular, therefore, by homogeneity and independence of increments, $m(t+s) = m(t) + m(s)$ and $v(t+s) = v(t) + v(s)$. The only continuous solutions are $m(t) = m(1)t$ and $v(t) = v(1)t$. Thus, let us focus on the form of the Lévy measure ν. For $\delta > 0$, let B_δ be a Borel subset of $\{z \in \mathbb{R} : |z| \geq \delta\}$. The process $N_\delta(t) := N((0,t] \times B_\delta), t \geq 0$, is a homogeneous Poisson process, and therefore $\nu_\delta(t) := \nu((0,t] \times B_\delta) = \mathbb{E}N_\delta(t)$ is continuous and linear in t, i.e., $\nu_\delta(t) = \nu_\delta(1)t, t \geq 0$. Let B be a Borel subset of $\mathbb{R}\backslash\{0\}$ and observe $B = \cup_{n=1}^\infty B_{\frac{1}{n}}$, where $B_{\frac{1}{n}} := B \cap \{z : |z| \geq \frac{1}{n}\}$ is a nondecreasing sequence of Borel subsets of $\mathbb{R}\backslash\{0\}$ bounded away from 0. Define $\tilde{\nu}(B) = \lim_{n\to\infty} \nu_{\frac{1}{n}}(1) = \lim_{n\to\infty} \nu((0,1] \times B_{\frac{1}{n}})$. Then, $B \to \tilde{\nu}(B)$ is a well-defined measure on the Borel subsets of $\mathbb{R}\backslash\{0\}$. Moreover, $\nu(dt \times dz) = \tilde{\nu}(dz) \times dt$, and therefore, $\int_{\mathbb{R}\backslash\{0\}} \frac{z^2}{1+z^2}\tilde{\nu}(dz) = \int\int_{\{(s,z):0<s\leq 1, z\neq 0\}} \frac{z^2}{1+z^2}\nu(ds \times dz) < \infty$. ■

Remark 5.7 The measure $\tilde{\nu}$ is usually referred to as the (homogeneous) Lévy measure in place of $\nu(dt \times dz) = dt \times \tilde{\nu}(dz)$ in the case of homogeneous Lévy processes. The use of the "tilde" notation is generally omitted, and the meaning of "Lévy measure" $\nu(dt \times dz)$ versus $\nu(dz)$ is left to be determined from the context.

Definition 5.7 A homogeneous Lévy process $X = \{X_t : t \geq 0\}$, such that for every $\lambda > 0$, there are unique constants $b_\lambda > 0$, $c_\lambda \in \mathbb{R}$ with the property that

$$\{X_{\lambda t} : t \geq 0\} =^d \{b_\lambda X_t + c_\lambda t : t \geq 0\}, \tag{5.39}$$

where $=^d$ denotes equality "in distribution" of the two respective rescaled processes, is referred to as a *(Lévy) stable process*.

The trivial case in which $X_t = ct, t \geq 0$, is referred to as the degenerate case of a stable process.

Corollary 5.21 *If $X = \{X_t : t \geq 0\}$ is a nondegenerate stable process, then there is a $0 < \theta \leq 2$ and $\nu_+ \geq 0$, $\nu_- \geq 0$, such that*

$$b_\lambda = \lambda^{\frac{1}{\theta}}, \quad \nu(dz) = \frac{\nu_+}{z^{1+\theta}} \mathbf{1}_{(0,\infty)}(z)dz + \frac{\nu_-}{|z|^{1+\theta}} \mathbf{1}_{(-\infty,0)}(z)dz.$$

Also, $\nu_+ + \nu_- > 0$ if and only if $\theta \in (0, 2)$. In particular, for $\theta \in (0, 2)$, $r = \int_{R_0} \frac{z}{1+z^2}$ exists and is finite and

$$\mathbb{E}e^{i\xi(X_t - X_s)} = \exp\{[i\xi(m + r) - \frac{v}{2}\xi^2 + \int_{\mathbb{R}_0} (e^{i\xi z} - 1)\nu(dz)](t - s)\}.$$

Proof For fixed $\lambda > 0$, note that each of the processes $X_\lambda^{(1)} := \{X_{\lambda t} : t \geq 0\}$ and $X_\lambda^{(2)} := \{b_\lambda X_t + c_\lambda t : t \geq 0\}$ is a homogeneous Lévy processes, since X is one. Let $N_\lambda^{(j)}$, $j = 1, 2$, denote their respective associated space-time jump counting measures, and N, the space-time jump counting measure associated with X. Then, for $z > 0, t > 0$,

$$\begin{aligned}
\mathbb{E}N_\lambda^{(1)}((0, t] \times (z, \infty)) &= \mathbb{E}\text{card}\{s \leq t : X_{\lambda s} - X_{\lambda s^-} > z\} \\
&= \mathbb{E}\text{card}\{s \leq \lambda t : X_s - X_{s^-} > z\} \\
&= \mathbb{E}N((0, \lambda t] \times (z, \infty)) = \lambda t \nu(z, \infty).
\end{aligned}$$

Similarly

$$\mathbb{E}N_\lambda^{(2)}((0, t] \times (z, \infty)) = \mathbb{E}N((0, t] \times (b_\lambda^{-1} z, \infty)) = t\nu(b_\lambda^{-1} z, \infty).$$

Since $X_\lambda^{(1)}$ and $X_\lambda^{(2)}$ have the same distribution, the respective space-time counting measures are equally distributed, and therefore,

$$\lambda \nu(z, \infty) = \nu(b_\lambda z, \infty) \quad \text{for all } z > 0, \lambda > 0.$$

Now for any two positive numbers λ_1, λ_2, and $z > 0$, one has

$$\nu(b^{-1}_{\lambda_1 \lambda_2} z, \infty) = \lambda_1 \lambda_2 \nu(z, \infty) = \nu(b^{-1}_{\lambda_1} b^{-1}_{\lambda_2} z, \infty).$$

Thus, if $\nu(1, \infty) > 0$, then $b_{\lambda_1 \lambda_2} = b_{\lambda_1} b_{\lambda_2}$ is positive, continuous, and log-linear on $(0, \infty)$. Hence, $b_\lambda = \lambda^\alpha$ for some real number α. Moreover, since $\lambda \nu(z, \infty)$ is nondecreasing in λ, one has $\alpha > 0$. Let $\theta = \frac{1}{\alpha} > 0$. Taking $\lambda = z^\theta$, one obtains $z^\theta \nu(z, \infty) = \nu(1, \infty)$ and hence, writing $\nu^+ = \nu(1, \infty) > 0$, we have

$$\nu(dz) = \frac{\nu^+}{z^{\theta+1}} dz \quad \text{on } (0, \infty).$$

Defining $\nu^+ = 0$ in the case $\nu(1, \infty) = 0$ makes the formula true when $\nu(1, \infty) = 0$ as well. By repeating all of the above considerations with $z < 0$, one obtains $\nu^- = \nu(-\infty, 0) \geq 0$ such that

$$\nu(dz) = \frac{\nu^-}{|z|^{\theta+1}} dz \quad \text{on } (-\infty, 0).$$

To check that one obtains the same exponent θ in both cases, use the relation $\mathbb{E} N^{(1)}_\lambda((0, t] \times (-\infty, -z) \cup (z, \infty)) = \mathbb{E} N^{(2)}_\lambda((0, t] \times (-\infty, -z) \cup (z, \infty))$. Finally, note that since $\int_{\{z:|z| \leq 1\}} z^2 \nu(dz) < \infty$ and $\int_{\{z:|z|>1\}} \nu(dz) < \infty$, one has either $\nu^+ = \nu^- = 0$ or $0 < \theta < 2$. In this case, the formula for the characteristic function follows directly from the Lévy-Khinchine formula. If $\nu^+ = \nu^- = 0$, then $N = 0$ and

$$X_t = mt + \sqrt{v} B_t, \quad (v > 0), t \geq 0.$$

The nondegeneracy assumption is used to obtain $v > 0$. Standard scaling properties of Brownian motion now imply $\theta = 2$. ■

Definition 5.8 The parameter θ is called the *stable law exponent*. In the case $c_\lambda = 0$, the stable process is said to be *self-similar*.

Example 3 provides an important self-similar example. The scaling relations (5.10) and (5.39) are easily seen to be equivalent by replacing λ by $\lambda^{-\frac{1}{\theta}}$ in (5.10), since $\alpha(\lambda) = \lambda^\theta$ in the latter.

Exercises

1. Show that if $X_t = \sum_{j=1}^{N_t} Y_j, t > 0$, $X_0 = 0$, is a compound Poisson process with i.i.d. nonnegative increments $Y_j, j \geq 1$, independent of the Poisson process $N_t, t \geq 0$ with intensity parameter $\rho > 0$, then $m = 0$ in Lemma 1, and $\nu(dx) = \rho P(Y_1 \in dx)$.
2. For the Poisson random measure, show that $\rho(A) = \mathbb{E}N(A)$, $A \in \mathcal{S}$, defines a measure and verify the formula (5.5).
3. Show that the counting process in $\{N(t) : t \geq 0\}$ in the proof of Corolllary 5.8 is continuous in probability.
4. Verify the multiplicative property (5.11). [*Hint*: Iteratively calculate $e^{-t\alpha(\lambda_1\lambda_2)\psi(1)} = \mathbb{E}e^{-X_{\alpha(\lambda_1\lambda_2)t}}$.]
5. *(Heavy Tails)* Suppose that X is a stable law with $\mathbb{E}|X|^m < \infty$ for all $m = 1, 2, \ldots$. Show that X must be Gaussian. If X has a stable law of exponent $\theta \in (0, 2)$, determine the largest real number m such that $\mathbb{E}|X|^m < \infty$.
6. *(Self-similarity)* A real-valued stochastic process $X = \{X_t : t \geq 0\}$, $X_0 = 0$, is said to be *self-similar* if for each $\lambda > 0$ there is a constant $c_\lambda > 0$ such that the stochastic process $\{X_{\lambda t} : t \geq 0\}$ has the same distribution as $\{c_\lambda X_t : t \geq 0\}$.

   (i) Show that standard Brownian motion started at 0 is self-similar with $c_\lambda = \lambda^{\frac{1}{2}}$.
   (ii) Show that in general $c_\lambda = \lambda^\gamma$ for some parameter $\gamma \geq 0$, referred to as the *scaling exponent*.
   (iii) Show that if $X = \{X_t : t \geq 0\}$ is a stochastically continuous, stationary, self-similar process, then X is a.s. constant. [*Hint*: Note that by stationarity $X_{t+h} = X_t$ in distribution for arbitrary $h \geq 0$.]

7. Show that the Gaussian process in Theorem 5.5 may be represented as a deterministic translation of a time change of Brownian motion.
8. Let $\lambda > 0$ and let τ_a be the first passage time to $a > 0$ for standard Brownian motion B starting at zero. Show that $\tau_a < \infty$ a.s., and $\mathbb{E}e^{-\lambda\tau_a} = e^{-\sqrt{2\lambda}a}$. [*Hint*: For $y > 0$, $P(\max_{0\leq s\leq t} B_s > y) \leq P(\max_{0\leq s\leq t}\{B_s - \frac{y}{2t}s\} > \frac{y}{2}) \leq e^{-\frac{y^2}{2t}}$, so that $P(\tau_a < \infty) = 1$. Let $M_t = \exp\{\lambda B_t - \frac{\lambda^2 t}{2}\}, t \geq 0$. Then, $1 = \mathbb{E}M_{t\wedge\tau_a}$ since $M_{t\wedge\tau_a}, t \geq 0$, is a martingale. Let $t \to \infty$.]
9. *(Infinitely Divisible Transition Probabilities)* Supppose that $X = \{X_t : t \geq 0\}$ is a homogeneous Lévy process. Apply the theory developed in this chapter to show that X is a Markov process on $\mathbb{R}$ with transition probabilities $p(t; x, B) = q_t(B - x), x \in \mathbb{R}, B \in \mathcal{B}(\mathbb{R})$, where q_t is a probability on $\mathbb{R}$ with characteristic function of the form $\hat{q}_t(\xi) = \exp(t\varphi(\xi))$, where $\varphi(\xi) = im\xi - \frac{v}{2}\xi^2 + \int_{\mathbb{R}}(e^{i\xi z} - 1 - \frac{i\xi z}{1+z^2})\nu(dz)$.
10. Let $\mathbf{B} = \{(B_t^{(1)}, B_t^{(2)}) : t \geq 0\}$ denote two-dimensional standard Brownian motion starting at $(0, 0)$. Let $\tau_y = \inf\{t \geq 0 : B_t^{(2)} \geq y\}$. Show that $C = \{C_y := B_{\tau_y}^{(1)} : y \geq 0\}$ is a (symmetric) Cauchy process, i.e., a Lévy

process whose increments are distributed according to the symmetric Cauchy distribution. [*Hint*: Use the strong Markov property to argue that $\{C_y : y \geq 0\}$ is a Markov process, and use the first passage time density for $\{B_t^{(2)}, t \geq 0\}$ and a change of variables to compute the stationary transition probabilities $p(y; x, dz) = \frac{1}{\pi} \frac{y}{y^2+(z-x)^2} dz$ for C.]

11. Show that for processes with right-continuous sample paths having left limits, continuity in probability at t is equivalent to the condition that $X_t = X_{t^-}$ a.s.
12. Let $X = \{X_t : t \geq 0\}$ and $Y = \{Y_t : t \geq 0\}$ be homogeneous Poisson processes defined on a probability space $(\Omega, \mathcal{F}, P)$. Suppose that a.s. X and Y do not have a common jump time. Give an alternative argument to Lemma 2 for this special case, to prove that X and Y are independent based on their respective holding time sequences.
13. Calculate the characteristic functions $\widehat{Q}_t(\xi) = \mathbb{E}e^{i\xi(X_t - X_0)}$ in the following. In each case, show directly that the distribution of X_t is infinitely divisible.

    (a) $\{X_t : t \geq 0\}$ is Brownian motion.
    (b) $\{X_t : t \geq 0\}$ is a compound Poisson process.

14. Suppose that $X_t = Y_t + Z_t, t \geq 0$, where $Y = \{Y_t : t \geq 0\}$ is a compound Poisson process and $Z = \{Z_t : t \geq 0\}$ is a Brownian motion, independent of Y. Show that $\{X_t : t \geq 0\}$ is a process with stationary independent increments.
15. Show that $J(\omega) := \{(t, X_t(\omega) - X_{t^-}(\omega)) : t \in \Gamma(\omega)\}$ is a countable set, where $\Gamma(\omega) := \{t > 0 : X_t(\omega) - X_{t^-}(\omega) \neq 0\}$. [*Hint*: Consider jumps of size $\frac{1}{n}$ or greater for each $n = 1, 2, \dots$ }.]
16. (i) Show that if $\{\mathcal{A}_{st}\}$ is additive, then the family of σ-field completions $\{\overline{\mathcal{A}_{st}}\}$ is also additive. (ii) Show that if X is adapted to an additive family $\{\mathcal{A}_{st}\}$, then X is an additive process and $\mathcal{D}_{st}(X) \subset \mathcal{A}_{st}$, provided that $X_0 = 0$ almost surely. (iii) Prove that additivity is preserved under convergence in probability.
17. Show that σ-field completions of an additive differential filtration yield an additive differential filtration. Show that $X = \{X_t : t \geq 0\}$, $X_0 = 0$, is an additive process if and only if it is adapted to an additive differential filtration.
18. Consider $\mathcal{R}_0 := \{A \in \mathcal{B}_0 : A \subset (0, \delta) \times \{\Delta : |\Delta| > \delta^{-1}\}$ for some $\delta > 0\}$. Show that $\mathcal{R}_0$ is closed under finite unions and relative complements, and countable intersections but generally not under countable unions nor complements. [*Hint*: If $\cap_n A_n = \emptyset$, then it belongs to $\mathcal{R}_0$ by definition. If $A \subset B \in \mathcal{R}_0$ then $B \backslash A \subset B$ implies $B \backslash A \in \mathcal{R}_0$.]
19. Define $N(A) = N(A, \omega) := \operatorname{card}\{A \cap J(\omega)\}, \quad A \in \mathcal{R}_0$. Show that $\omega \mapsto N(A, \omega)$ is measurable by completing the following steps: Let $\mathcal{D} = \{\mathcal{D}_{s,t} : 0 \leq s < t < \infty\}$ denote the additive differential filtration associated with the Lévy process X. For $0 < s < t, z > 0$, $S_{s,t,z} := (s, t] \times (z, \infty)$, and let $Q_{s,t}$ denote the rationals in $(s, t]$.

(i) Show $N(S_{s,t,z})$ is measurable with respect to $\{\overline{\mathcal{D}}_{s,t} : 0 < s < t < \infty\}$. [*Hint*: Show $[N(S_{s,t,z}) \geq 1] \in \overline{\mathcal{D}}_{s,t}$ by expressing as countable unions and intersections of events of the form $[X_q - X_p \geq z + 1/m]$, $p, q \in Q_{s,t}$. Then, proceed inductively by expressing $[N(S_{s,t,z}) \geq k+1] = \cup_{q \in Q_{s,t}}[N(S_{s,q,z}) \geq k] \cap [N(S_{q,t,z}) \geq 1]$.]

(ii) Show that $N(A \cap S_{s,t,z})$ is measurable with respect to $\overline{\mathcal{D}}_{s,t}$ for $A \in \mathcal{B}_0$. [*Hint*: Consider the collection of all $A \in \mathcal{B}_0$ such that $N(A \cap S_{s,t,z})$ is measurable with respect to $\overline{\mathcal{D}}_{s,t}$, and use the π-λ theorem.]

(iii) Show that $N(A \cap \tilde{S}_{s,t,z})$ is measurable with respect to $\overline{\mathcal{D}}_{s,t}$ for $A \in \mathcal{B}_0$, where $\tilde{S}_{s,z,t} := (s,t] \times (-\infty, -z)$, $z > 0$. [*Hint*: Repeat similar steps as for $S_{s,t,z}$.]

(iv) For $A \in \mathcal{R}_0$, show $N(A)$ is measurable with respect to $\overline{\mathcal{D}}_{s,t}$. [*Hint*: Consider $A \in \mathcal{R}_0$ with $A \subset (s,t] \times \mathbb{R}\backslash\{0\}$ for some $0 < s < t$. Consider $A = A \cap S_{s,t,z} \cup A \cap \tilde{S}_{s,t,z}$ for small $z > 0$.]

20. Show that if one also assumes stationarity of increments in Theorem 5.7, then one may in fact directly conclude that $\rho_n = np_n \leq -\log P(X_t = 0)$ and $P(X_t = 0) > 0$. [*Hint*: $P(X_t = 0) = P(\cap_{i=1}^n A_i^{(n)^c}) = (1-p_n)^n$. Conclude that if $P(X_t = 0) = 0$, then $P(A_i^{(n)}) = 1$ for each i and hence $X_t \geq n$ for each n.]

Chapter 6
The Stochastic Integral

In this chapter, Itô's stochastic integral and some of its basic properties are carefully introduced. The essential role of martingale theory is made manifestly clear through this development.

A one-dimensional Brownian motion $X = \{X_t : t \geq 0\}$ starting at x and having drift coefficient μ and diffusion coefficient σ^2 may be defined by the transformation of standard Brownian motion $B = \{B_t : t \geq 0\}$ given by

$$X_t = x + \mu t + \sigma B_t, \quad t \geq 0. \tag{6.1}$$

Symbolically, this may be viewed as the integration, which "defines" the equation in differential form denoted by

$$dX_t = \mu dt + \sigma dB_t, \quad X_0 = x. \tag{6.2}$$

Along these same lines, a diffusion $X = \{X_t : t \geq 0\}$ on $\mathbb{R}$ may be thought of as a Markov process that is locally like a Brownian motion. That is, in a sense to be made precise, the following relation holds,

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t, \tag{6.3}$$

where $\mu(\cdot), \sigma(\cdot)$ are given functions on $\mathbb{R}$ and $B = \{B_t : t \geq 0\}$ is a standard one-dimensional Brownian motion. In other words, conditionally given $\{X_s : 0 \leq s \leq t\}$, for sufficiently smooth $\mu(\cdot)$ and $\sigma^2(\cdot)$, in a small time interval $(t, t + dt]$, the displacement $dX_t := X_{t+dt} - X_t$ is approximately the Gaussian random variable

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_6

$\mu(X_t)dt + \sigma(X_t)(B_{t+dt} - B_t)$, having mean $\mu(X_t)dt$ and variance $\sigma^2(X_t)dt$. However, the sample paths of Brownian motion are continuous yet wild (nowhere differentiable). To be precise, one has the following result.

Proposition 6.1 *Let* $\mathbf{B} = \{B_t : t \geq 0\}$ *be a standard Brownian motion starting at zero. Fix* $t > 0$. *Then, one has*

$$\max_{1\leq k\leq 2^n} |\sum_{m=0}^{k-1}(B_{(m+1)2^{-n}t} - B_{m2^{-n}t})^2 - k2^{-n}t| \to 0 \quad \text{a.s.} \qquad \text{as } n \to \infty. \tag{6.4}$$

Proof The partial sums $Y_k \equiv Y_{k,n} = \sum_{m=0}^{k-1}(B_{(m+1)2^{-n}t} - B_{m2^{-n}t})^2 - k2^{-n}t$, $k = 1, \ldots, 2^n$, form a martingale since the summands are mean zero independent random variables. Writing $N = 2^n$ and noting that the N terms are i.i.d. with the common distribution of $Z^2 - \mathbb{E}Z^2$, where Z is $N(0, 2^{-n}t)$, it follows from Doob's maximal inequality (Corollary 1.16) that for all n,

$$\begin{aligned} P(\max_{1\leq k\leq 2^n} |Y_k| \geq 2^{-\frac{n}{4}}) &\leq \mathbb{E}Y_N^2/2^{-\frac{n}{2}} \\ &= N2^{\frac{n}{2}} var(Z^2) = 2^{\frac{3}{2}n} var(Z^2) \\ &\leq 2^{\frac{3}{2}n}\mathbb{E}Z^4 = 2^{\frac{3}{2}n}3(\frac{t}{N})^2 = 3t^2 2^{-\frac{n}{2}}. \end{aligned} \tag{6.5}$$

Since the right-hand side is summable (over $n = 1, 2, \ldots$), it follows from the Borel-Cantelli lemma that there is a finite $n_0(\omega)$, depending on the sample point $\omega \in \Omega$, such that

$$P(\max_{1\leq k\leq 2^n} |Y_k| < 2^{-\frac{n}{4}} \text{ for all } n \geq n_0) = 1. \tag{6.6}$$

In particular, this implies (6.4). ■

As a result of Proposition 6.1, outside a set of probability 0, a Brownian path is of unbounded variation on every nondegenerate interval (Exercise 1). Thus, (6.3) *cannot* be regarded as an ordinary differential equation such as

$$dX_t/dt = \mu(X_t) + \sigma(X_t)dB_t/dt.$$

The precise sense in which (6.3) is valid and, in fact, may be solved to construct a diffusion with coefficients $\mu(\cdot), \sigma^2(\cdot)$, is the goal of this chapter and the next.

The first task is to provide a useful definition of the integrated version of (6.3) expressed as:

$$X_t = x + \int_0^t \mu(X_s)ds + \int_0^t \sigma(X_s)dB_s. \tag{6.7}$$

The first integral may generally be viewed as an ordinary Riemann integral in cases, for example, when $\mu(\cdot)$ is continuous and $s \to X_s$ is continuous. But the second integral cannot be defined as a Riemann-Stieltjes integral, since the Brownian paths are of unbounded variation on $[0, t]$ for every $t > 0$. On the other hand, for a constant function $\sigma(x) \equiv \sigma$, the second integral has the obvious meaning as $\sigma \cdot (B_t - B_0)$, so that (6.7) becomes

$$X_t = x + \int_0^t \mu(X_s)ds + \sigma(B_t - B_0). \tag{6.8}$$

It turns out that if, for example, $\mu(x)$ is Lipschitzian, then (6.8) may be solved sample path-wise, more or less by Picard's well-known method of iteration for ordinary differential equations, to obtain a continuous solution $t \to X_t$; see Exercise 3.

To set the general mathematical framework for defining the stochastic integral, let $(\Omega, \mathcal{F}, P)$ be a probability space on which a standard one-dimensional Brownian motion $B = \{B_t : t \geq 0\}$ is defined. Suppose $\{\mathcal{F}_t : t \geq 0\}$ is an increasing family of sub-σ-fields of $\mathcal{F}$, referred to as a *filtration* of $\mathcal{F}$, such that

$$\begin{aligned} &\text{(i) } B_s \text{ is } \mathcal{F}_s\text{-measurable } (s \geq 0).\\ &\text{(ii) } \{B_t - B_s : t \geq s\} \text{ is independent of (events in) } \mathcal{F}_s (s \geq 0). \end{aligned} \tag{6.9}$$

Condition (i) says that the Brownian motion $B = \{B_t : t \geq 0\}$ is *adapted* to the filtration $\{\mathcal{F}_t : t \geq 0\}$. For example, one may take $\mathcal{F}_t = \sigma\{B_s : 0 \leq s \leq t\}$. This is the smallest $\mathcal{F}_t$ one may take in view of (i). Often it is important to take larger $\mathcal{F}_t$. As an example, let $\mathcal{F}_t = \sigma[\{B_s : 0 \leq s \leq t\}, \{Z_\lambda : \lambda \in \Lambda\}]$, where $\{Z_\lambda : \lambda \in \Lambda\}$ is a family of random variables independent of $\{B_t : t \geq 0\}$. For technical convenience, also assume that $\mathcal{F}_t, \mathcal{F}$ are *P–complete*, i.e., if $N \in \mathcal{F}$ and $P(N) = 0$, then N and each of its subsets belong toeach $\mathcal{F}_t$.

For the ensuing definitions, fix two numbers (time points) $0 \leq \alpha < \beta < \infty$.

Definition 6.1 A process $f = \{f(t) : \alpha \leq t \leq \beta\}$ is *progressively measurable* if, for each t, the map $(s, \omega) \to f(s, \omega)$ is $\mathcal{B}[\alpha, t] \otimes \mathcal{F}_t$-measurable on $[\alpha, t] \times \Omega$, where $\mathcal{B}[\alpha, t]$ is the Borel σ-field on $[\alpha, t]$.

Remark 6.1 If the process $f = \{f_t : \alpha \leq t \leq \beta\}$ is either right-continuous, or left-continuous, then it is progressively measurable (e.g., see BCPT,[1] Proposition 3.6, p. 60.).

Definition 6.2 A real-valued progressively measurable stochastic process $f = \{f(t) : \alpha \leq t \leq \beta\}$ is said to be a *nonanticipative functional on* $[\alpha, \beta]$. Such an f is said to *belong to* $\mathcal{M}[\alpha, \beta]$ if

[1] Throughout, BCPT refers to Bhattacharya and Waymire (2016), A Basic Course in Probability Theory.

$$\mathbb{E}\int_{\alpha}^{\beta} f^2(t)dt < \infty. \tag{6.10}$$

If $f \in \mathcal{M}[\alpha, \beta]$ for all $\beta > \alpha$, then f is said to *belong to* $\mathcal{M}[\alpha, \infty)$.

Remark 6.2 It is worthwhile for later considerations to note that $\mathcal{M}[\alpha, \beta]$ is a subspace of $L^2([\alpha, \beta] \times \Omega, \mathcal{B}[\alpha, \beta] \otimes \mathcal{F}, \lambda \times P)$ with the norm

$$||f||^2_{L^2([\alpha,\beta]\times\Omega,\mathcal{B}\otimes\mathcal{F},\lambda\times P)} := \mathbb{E}(\int_{\alpha}^{\beta} |f(s)|^2 ds), \quad f \in \mathcal{M}([\alpha, \beta]). \tag{6.11}$$

To initiate the definition of the stochastic integral, one may begin by consideration of nonanticipative random step functionals to serve as integrands in a type of "Riemann sum". One may easily check by inspection that the following step functionals are indeed non-anticipative.

Definition 6.3 A real-valued stochastic process $\{f(t) : \alpha \le t \le \beta\}$ is said to be a *nonanticipative step functional on* $[\alpha, \beta]$ if there exists a finite set of time points $\alpha = t_0 < t_1 < \cdots < t_m = \beta$ and random variables f_j $(0 \le j \le m)$, such that

(i) f_j is $\mathcal{F}_{t_j}$-measurable $(0 \le j \le m)$.

(ii) $f(t) = f_j$ for $t_j \le t < t_{j+1}$ $(0 \le j \le m-1)$, $f(\beta) = f_m$. (6.12)

Definition 6.4 (Itô Integral of a Step Functional) The *stochastic integral*, or the *Itô integral*, of the nonanticipative step functional $f = \{f(t) : \alpha \le t \le \beta\}$ in (6.12) is the stochastic process defined for $\alpha \le t \le \beta$ by

$$\int_{\alpha}^{t} f(s)dB_s$$
$$:= \begin{cases} f(t_0)(B_t - B_{t_0}) \equiv f(\alpha)(B_t - B_{\alpha}) \text{ for } t \in [\alpha, t_1], \\ \sum_{j=1}^{k} f(t_{j-1})(B_{t_j} - B_{t_{j-1}}) + f(t_k)(B_t - B_{t_k}) \text{ for } t \in (t_k, t_{k+1}]. \end{cases} \tag{6.13}$$

Observe that the Riemann-type sum (6.13) is obtained by taking the value of the integrand at the *left end point* of a time interval $(t_{j-1}, t_j]$. As a consequence, for each $t \in [\alpha, \beta]$ the Itô integral $\int_{\alpha}^{t} f(s)dB_s$ is $\mathcal{F}_t$*–measurable*, i.e., it is nonanticipative.

Proposition 6.2 *The Itô integral given in Definition 6.4 has the following properties.*

a. *(Continuous, Nonanticipative, Additive) If f is a nonanticipative step functional on $[\alpha, \beta]$, then $t \to \int_{\alpha}^{t} f(s)dB_s$ is nonanticipative and is continuous for every $\omega \in \Omega$. The integral is also additive in the sense*

$$\int_{\alpha}^{s} f(u)dB_u + \int_{s}^{t} f(u)dB_u = \int_{\alpha}^{t} f(u)dB_u, \quad (\alpha \le s < t \le \beta). \tag{6.14}$$

b. *(Linearity) If f, g are nonanticipative step functionals on $[\alpha, \beta]$, then for any $a, b \in \mathbb{R}$, $af + bg$ is a nonanticipative step functional and*

$$\int_{\alpha}^{t} (af(s) + bg(s))dB_s = a\int_{\alpha}^{t} f(s)dB_s + b\int_{\alpha}^{t} g(s)dB_s. \tag{6.15}$$

c. *(Martingale) Suppose f is a nonanticipative step functional on $[\alpha, \beta]$ such that*

$$\mathbb{E}\int_{\alpha}^{\beta} f^2(t)dt \equiv \int_{\alpha}^{\beta} (\mathbb{E}f^2(t))dt < \infty. \tag{6.16}$$

Then, $\{\mathcal{I}_\alpha(t) := \int_\alpha^t f(s)dB_s : \alpha \le t \le \beta\}$ is a square integrable $\{\mathcal{F}_t\}$–martingale, i.e., $\mathbb{E}(\int_\alpha^t f(s)dB_s)^2 < \infty$ and the martingale property,

$$\mathbb{E}(\mathcal{I}_\alpha(t) \mid \mathcal{F}_s) = \mathbb{E}(\int_{\alpha}^{t} f(u)dB_u \mid \mathcal{F}_s) = \int_{\alpha}^{s} f(u)dB_u = \mathcal{I}_\alpha(s), \tag{6.17}$$

holds for $\alpha \le s < t \le \beta$.

d. *(Centered) If (6.16) holds, then*

$$\mathbb{E}(\int_{\alpha}^{t} f(s)dB_s|\mathcal{F}_\alpha) = 0, \quad \mathbb{E}\int_{\alpha}^{t} f(s)dB_s = 0 \tag{6.18}$$

e. *(Itô Isometry—L^2 Norm Preserving Property) If f, g are square-integrable, nonanticipative step functionals on $[\alpha, \beta]$, then*

$$\mathbb{E}((\int_{\alpha}^{t} f(s)dB_s)^2|\mathcal{F}_\alpha) = \mathbb{E}(\int_{\alpha}^{t} f^2(s)ds \mid \mathcal{F}_\alpha)$$

and, more generally,

$$\mathbb{E}((\int_{\alpha}^{t} f(s)dB_s)(\int_{\alpha}^{t} g(s)dB_s)|\mathcal{F}_\alpha) = \mathbb{E}(\int_{\alpha}^{t} f(s)g(s)ds \mid \mathcal{F}_\alpha) \tag{6.19}$$

for $\alpha \le t \le \beta$. In particular,

$$\begin{aligned} < \int_{\alpha}^{t} f(s)dB_s, \int_{\alpha}^{t} g(s)dB_s >_{L^2(\Omega,P)} &\equiv \mathbb{E}(\int_{\alpha}^{t} f(s)dB_s)(\int_{\alpha}^{t} g(s)dB_s) \\ &= \mathbb{E}\int_{\alpha}^{t} f(s)g(s)ds \\ &:= \langle f, g\rangle_{L^2([\alpha,t]\times\Omega,\lambda\times P)}. \end{aligned} \tag{6.20}$$

f. *(Quadratic Martingales) Suppose that f is a square-integrable, nonanticipative functional. The process*

$$Q_\alpha(t) := (\int_\alpha^t f(s)dB_s)^2 - \int_\alpha^t f^2(s)ds, \quad (a \le t \le \beta), \tag{6.21}$$

is a martingale.

Proof (a) and (b) follow from Definition 6.4.

(c) Let f be given by (6.12). As $\int_\alpha^s f(u)dB_u$ is $\mathcal{F}_s$-measurable, it follows from (6.14) that

$$\mathbb{E}(\int_\alpha^t f(u)dB_u \mid \mathcal{F}_s) = \int_\alpha^s f(u)dB_u + \mathbb{E}(\int_s^t f(u)dB_u \mid \mathcal{F}_s). \tag{6.22}$$

Now, by (6.13), if $t_{j-1} < s \le t_j$ and $t_i < t \le t_{i+1}$, then

$$\begin{aligned}\int_s^t f(u)dB_u = f(s)(B_{t_j} - B_s) + f(t_j)(B_{t_{j+1}} - B_{t_j})\\ + \cdots + f(t_{i-1})(B_{t_i} - B_{t_{i-1}}) + f(t_i)(B_t - B_{t_i}).\end{aligned} \tag{6.23}$$

Observe that, for $s' < t'$, $B_{t'} - B_{s'}$ is independent of $\mathcal{F}_{s'}$ (property (6.9)(ii)), so that

$$\mathbb{E}(B_{t'} - B_{s'} \mid \mathcal{F}_{s'}) = 0 \qquad (s' < t'). \tag{6.24}$$

Applying this to (6.23),

$$\begin{aligned}\mathbb{E}(f(s)(B_{t_j} - B_s) \mid \mathcal{F}_s) &= f(s)\mathbb{E}(B_{t_j} - B_s \mid \mathcal{F}_s) = 0,\\ \mathbb{E}(f(t_j)(B_{t_{j+1}} - B_{t_j}) \mid \mathcal{F}_s) &= \mathbb{E}[\mathbb{E}(f(t_j)(B_{t_{j+1}} - B_{t_j}) \mid \mathcal{F}_{t_j}) \mid \mathcal{F}_s]\\ &= \mathbb{E}[f(t_j)\mathbb{E}(B_{t_{j+1}} - B_{t_j} \mid \mathcal{F}_{t_j}) \mid \mathcal{F}_s] = 0\\ \mathbb{E}(f(t_i)(B_t - B_{t_i}) \mid \mathcal{F}_s) &= \mathbb{E}[f(t_i)\mathbb{E}(B_t - B_{t_i}) \mid \mathcal{F}_s)] = 0.\end{aligned} \tag{6.25}$$

Therefore,

$$\mathbb{E}(\int_s^t f(u)dB_u \mid \mathcal{F}_s) = 0 \qquad (s < t). \tag{6.26}$$

From this and (6.14), the martingale property (6.17) follows. The relation (6.18) is an immediate consequence of (6.26). In order to prove the first inequality in (e), let $t \in [\alpha, t_1]$. Then,

$$\begin{aligned}\mathbb{E}((\int_\alpha^t f(s)dB_s)^2|\mathcal{F}_\alpha) &= \mathbb{E}(f^2(\alpha)(B_t - B_\alpha)^2 \mid \mathcal{F}_\alpha) \\ &= f^2(\alpha)\mathbb{E}((B_t - B_\alpha)^2 \mid \mathcal{F}_\alpha) \\ &= f^2(\alpha)(t-\alpha) = \mathbb{E}(\int_\alpha^t f^2(s)ds \mid \mathcal{F}_\alpha), \end{aligned} \tag{6.27}$$

by independence of $\mathcal{F}_\alpha$ and $B_t - B_\alpha$. If $t \in (t_i, t_{i+1}]$, then, by (6.13),

$$\begin{aligned}&\mathbb{E}((\int_\alpha^t f(s)dB_s)^2|\mathcal{F}_\alpha) \\ &= \mathbb{E}[\{\sum_{j=1}^{i} f(t_{j-1})(B_{t_j} - B_{t_{j-1}}) + f(t_i)(B_t - B_{t_i})\}^2|\mathcal{F}_\alpha]. \end{aligned} \tag{6.28}$$

Now the contribution of the product terms in (6.28) is zero. For, if $j < k$, then

$$\begin{aligned}&\mathbb{E}[f(t_{j-1})(B_{t_j} - B_{t_{j-1}})f(t_{k-1})(B_{t_k} - B_{t_{k-1}}) \mid \mathcal{F}_\alpha] \\ &= \mathbb{E}[\mathbb{E}(\cdots \mid \mathcal{F}_{t_{k-1}}) \mid \mathcal{F}_\alpha] \\ &= \mathbb{E}[f(t_{j-1})(B_{t_j} - B_{t_{j-1}})f(t_{k-1})\mathbb{E}(B_{t_k} - B_{t_{k-1}} \mid \mathcal{F}_{t_{k-1}}) \mid \mathcal{F}_\alpha] \\ &= 0. \end{aligned} \tag{6.29}$$

Therefore, (6.28) reduces to

$$\begin{aligned}&\mathbb{E}[(\int_\alpha^t f(s)dB_s)^2|\mathcal{F}_\alpha] \\ &= \sum_{j=1}^{i} \mathbb{E}(f^2(t_{j-1})(B_{t_j} - B_{t_{j-1}})^2 \mid \mathcal{F}_\alpha) + \mathbb{E}(f^2(t_i)(B_t - B_{t_i})^2 \mid \mathcal{F}_\alpha) \\ &= \sum_{j=1}^{i} \mathbb{E}[f^2(t_{j-1})\mathbb{E}((B_{t_j} - B_{t_{j-1}})^2 \mid \mathcal{F}_{t_{j-1}}) \mid \mathcal{F}_\alpha] \\ &\quad + \mathbb{E}[f^2(t_i)\mathbb{E}((B_t - B_{t_i})^2 \mid \mathcal{F}_{t_i}) \mid \mathcal{F}_\alpha] \\ &= \sum_{j=1}^{i} \mathbb{E}(f^2(t_{j-1})(t_j - t_{j-1}) \mid \mathcal{F}_\alpha) + \mathbb{E}(f^2(t_i)(t - t_i) \mid \mathcal{F}_\alpha) \\ &= \mathbb{E}(\sum_{j=1}^{i}(t_j - t_{j-1})f^2(t_{j-1}) + (t - t_i)f^2(t_i) \mid \mathcal{F}_\alpha) \\ &= \mathbb{E}(\int_\alpha^t f^2(s)ds \mid \mathcal{F}_\alpha). \end{aligned} \tag{6.30}$$

The relation (6.19) follows by the polar identity with respect to an innerproduct,

$$\langle f, g\rangle = \frac{1}{4}(||f+g||^2 - ||f-g||^2),$$

applied conditionally given $\mathcal{F}_\alpha$.

The last statement in (e) follows from the relation (6.19) by taking expectation. More general versions of (6.19) and (6.18) are proved later without appeal to the polar identity; see Theorem 7.3. The proof of (f) follows similarly from additivity of the integral and the Itô isometry property (e), using basic properties of the filtration $\{\mathcal{F}_t : \alpha \le t \le \beta\}$ and conditional expectation. ■

Remark 6.3 This proposition and its extension to more general class of integrands provides important classes of martingales, namely, Itô integrals and suitably centered squares of Itô integrals. Taking $f \equiv 1$ provides the special quadratic martingale $B_t^2 - t, t \ge 0$, already used, for example, to obtain the expected time to escape an interval for Brownian motion by optional stopping theory (Theorem 1.12). Specifically, letting

$$\tau_{a,b} = \inf\{t : B_t \in \{a, b\}\}, B_0 = x \in (a, b),$$

Theorem 1.12 may be applied to obtain

$$x = \mathbb{E}B_0 = \mathbb{E}B_{\tau_{a,b}} = aP(\tau_a \le \tau_b) + bP(\tau_a > \tau_b) = b + (a-b)P(\tau_a \le \tau_b), \tag{6.31}$$

and

$$x^2 = \mathbb{E}B_0^2 - 0 = \mathbb{E}B_\tau^2 - \mathbb{E}\tau, \tag{6.32}$$

from which it follows that

$$P(\tau_a \le \tau_b) = \frac{b-x}{b-a} \text{ and } \mathbb{E}\tau = (b-x)(x-a). \tag{6.33}$$

The process $\int_\alpha^t f^2(s)ds, \alpha \le t \le \beta$, is referred to as the *quadratic variation* of the martingale $\mathcal{I}_\alpha(t) = \int_\alpha^t f(s)dB_s, \alpha \le t \le \beta$. In particular, the standard Brownian motion $B = \{B_t : t \ge 0\}$ is a continuous martingale with quadratic variation t. The converse is true as well! This is Lévy's characterization of Brownian motion (see Exercise 1 in Chapter 8). Observe that according to property (f), the sub-martingale $(\int_\alpha^t f(s)dB_s)^2, \alpha \le t \le \beta$, differs from a martingale by the quadratic variation process $\langle f\rangle_t := \int_\alpha^t f^2(s)ds$. The connection between this definition of quadratic variation from the point of view of martingales and that familiar to analysis is developed in Exercise 2.

Remark 6.4 The proof of Proposition 6.2 suggests an often used heuristic formalism of stochastic calculus:

$$\mathbb{E}dB_t = \mathbb{E}(B_{t+dt} - B_t) = 0, \quad \mathrm{Cov}(dB_t, dB_s) = 0, s \neq t, \mathrm{Var}(dB_t) = dt. \tag{6.34}$$

We will see that the extension of the class of integrands to non-anticipative functionals in $\mathcal{M}[\alpha, \beta]$ is also consistent with this formalism.

The next task is to extend the definition of the stochastic integral to the larger class of nonanticipative functionals to serve as integrands, by approximating these by step functionals. The following is a standard result from real analysis, whose proof is included for completeness.

Proposition 6.3 *Let $f \in \mathcal{M}[\alpha, \beta]$. Then, there exists a sequence $\{f_n\}_{n=1}^{\infty}$ of nonanticipative step functionals belonging to $\mathcal{M}[\alpha, \beta]$ such that*

$$\mathbb{E}\int_{\alpha}^{\beta}(f_n(t) - f(t))^2 dt \longrightarrow 0 \quad \text{as } n \to \infty. \tag{6.35}$$

Proof The proof follows by a standard "molifier" argument from analysis for L^2-spaces. Extend $f(t)$ to $(-\infty, \infty)$ by setting it zero outside $[\alpha, \beta]$. For $\varepsilon > 0$, write $g_{\varepsilon}(t) := f(t - \varepsilon)$. Let ψ be a symmetric, continuous, nonnegative (nonrandom) function that vanishes outside $(-1, 1)$ and satisfies $\int_{-1}^{1}\psi(x)dx = 1$. Write $\psi_{\varepsilon}(x) := \psi(x/\varepsilon)/\varepsilon$. Then, ψ_{ε} vanishes outside $(-\varepsilon, \varepsilon)$ and satisfies $\int_{-\varepsilon}^{\varepsilon}\psi_{\varepsilon}(x)dx = 1$. Now define $g^{\varepsilon} := g_{\varepsilon} * \psi_{\varepsilon}$, i.e.,

$$\begin{aligned} g^{\varepsilon}(t) &= \int_{-\varepsilon}^{\varepsilon} g_{\varepsilon}(t-x)\psi_{\varepsilon}(x)dx = \int_{-\varepsilon}^{\varepsilon} f(t-\varepsilon-x)\psi_{\varepsilon}(x)dx \\ &= \int_{t-2\varepsilon}^{t} f(y)\psi_{\varepsilon}(t-\varepsilon-y)dy \quad (y = t-\varepsilon-x). \end{aligned} \tag{6.36}$$

Note that, for each $\omega \in \Omega$, the Fourier transform of g^{ε} is

$$\hat{g}^{\varepsilon}(\xi) = \hat{g}_{\varepsilon}(\xi)\hat{\psi}_{\varepsilon}(\xi) = e^{i\xi\varepsilon}\hat{f}(\xi)\hat{\psi}_{\varepsilon}(\xi),$$

so that, by the Plancherel identity (BCPT, (6.36), p. 112.),

$$\begin{aligned} \int_{-\infty}^{\infty}|g^{\varepsilon}(t) - f(t)|^2 dt &= \frac{1}{2\pi}\int_{-\infty}^{\infty}|\hat{g}^{\varepsilon}(\xi) - \hat{f}(\xi)|^2 d\xi \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}|\hat{f}(\xi)^2| \cdot |e^{i\xi\varepsilon}\hat{\psi}_{\varepsilon}(\xi) - 1|^2 d\xi. \end{aligned} \tag{6.37}$$

The last integrand is bounded by $4|\hat{f}(\xi)|^2$. Since

$$\mathbb{E}\int |\hat{f}(\xi)|^2 d\xi \equiv \int_{\Omega}\int_{-\infty}^{\infty} |\hat{f}(\xi)|^2 d\xi \, dP = 2\pi \int_{\Omega}\int_{-\infty}^{\infty} f^2(t) dt \, dP$$
$$= 2\pi \mathbb{E}\int_{[\alpha,\beta]} f^2(t) dt < \infty, \tag{6.38}$$

it follows, by Lebesgue's dominated convergence theorem, that

$$\mathbb{E}\int_{-\infty}^{\infty} |g^{\varepsilon}(t) - f(t)|^2 dt \longrightarrow 0 \quad \text{as } \varepsilon \to 0. \tag{6.39}$$

This proves that there exists a sequence $\{g_n\}_{n=1}^{\infty}$ of continuous nonanticipative functionals in $\mathcal{M}[\alpha, \beta]$ such that

$$\mathbb{E}\int_{\alpha}^{\beta} (g_n(s) - f(s))^2 ds \longrightarrow 0. \tag{6.40}$$

It is now enough to prove that if g is a continuous element of $\mathcal{M}[\alpha, \beta]$, then there exists a sequence $\{h_n\}_{n=1}^{\infty}$ of nonanticipative step functionals in $\mathcal{M}[\alpha, \beta]$ such that

$$\mathbb{E}\int_{\alpha}^{\beta} (h_n(s) - g(s))^2 ds \longrightarrow 0. \tag{6.41}$$

For this, first assume that g is also bounded, $|g| \leq M$. Define, for $0 \leq k \leq n-1$,

$$h_n(t) = g(\alpha + \frac{k}{n}(\beta - \alpha)) \text{ if } \alpha + \frac{k}{n}(\beta - \alpha) \leq t < \alpha + \frac{k+1}{n}(\beta - \alpha),$$
$$h_n(\beta) = g(\beta).$$

Use Lebesgue's dominated convergence theorem to prove (6.41). If g is unbounded, then apply (6.41) with g replaced by its truncation g^M, which agrees with g on $\{t \in [\alpha, \beta] : |g(t)| \leq M\}$ and equals $-M$ on the set $\{t \in [\alpha, \beta] : g(t) < -M\}$ and M on $\{t \in [\alpha, \beta] : g(t) > M\}$. Note that $g^M = (g \wedge M) \vee (-M)$ is a continuous and bounded element of $\mathcal{M}[\alpha, \beta]$, and apply Lebesgue's dominated convergence theorem to get

$$\mathbb{E}\int_{\alpha}^{\beta} (g^M(s) - g(s))^2 ds \longrightarrow 0 \quad \text{as } M \to \infty. \qquad \blacksquare$$

We are now ready to define the stochastic integral of an arbitrary element f in $\mathcal{M}[\alpha, \beta]$.

Theorem 6.4 *Given $f \in \mathcal{M}[\alpha, \beta]$, there is a sequence $\{f_n\}_{n=1}^{\infty}$ of square-integrable nonanticipative step functionals satisfying (6.35) and such that the sequence of stochastic processes defined by*

$$I_n(t) := \int_\alpha^t f_n(s)dB_s, \quad \alpha \le t \le \beta,$$

is a.s. uniformly Cauchy. That is, with probability one,

$$\max_{\alpha \le t \le \beta} |I_n(t) - I_m(t)| \to 0$$

as $n, m \to \infty$.

Proof Given $f \in \mathcal{M}[\alpha, \beta]$, let $\{f_n\}_{n=1}^\infty$ be a sequence of step functionals satisfying (6.35). For each pair of positive integers n, m, the stochastic integral $\int_\alpha^t (f_n - f_m)(s)dB_s$ $(\alpha \le t \le \beta)$ is a square integrable martingale, by Proposition 6.2(c). Therefore, by Doob's maximal inequality (Theorem 1.14) and the Itô isometry (6.19), for all $\varepsilon > 0$,

$$P(M_{n,m} > \varepsilon) \le \int_\alpha^\beta \mathbb{E}(f_n(s) - f_m(s))^2 ds/\varepsilon^2, \tag{6.42}$$

where

$$M_{n,m} := \max\{|\int_\alpha^t (f_n(s) - f_m(s))dB_s| : \alpha \le t \le \beta\}. \tag{6.43}$$

Choose an increasing sequence $\{n_k\}$ of positive integers such that the right side of (6.42) is less than $1/k^2$ for $\varepsilon = 1/k^2$, if $n, m \ge n_k$. Denote by A the set

$$A := \{\omega \in \Omega : M_{n_k, n_{k+1}}(\omega) \le 1/k^2 \text{ for all sufficiently large } k\}. \tag{6.44}$$

By the Borel-Cantelli lemma, $P(A) = 1$. Now on A the series $\sum_k M_{n_k, n_{k+1}} < \infty$. Thus, given any $\delta > 0$, there exists a positive integer $k(\delta)$ (depending on $\omega \in A$) such that

$$M_{n_j, n_\ell} \le \sum_{k \ge k(\delta)} M_{n_k, n_{k+1}} < \delta \qquad \forall\, j, \ell \ge k(\delta).$$

In other words, on the set A the sequence $\{\int_\alpha^t f_{n_k}(s)dB_s : \alpha \le t \le \beta\}$ is Cauchy in the supremum distance (6.43). Therefore, $\int_\alpha^t f_{n_k}(s)dB_s$ $(\alpha \le t \le \beta)$ converges uniformly to a continuous function, outside a set of probability zero. Simply relabel this subsequence as $\{f_n\}_{n=1}^\infty$ to obtain the assertion of the theorem. ■

Definition 6.5 For $f \in \mathcal{M}[\alpha, \beta]$, the uniform limit of $\int_\alpha^t f_n(s)dB_s$ $(\alpha \le t \le \beta)$ on A as constructed in Theorem 6.4 is called the *stochastic integral*, or the *Itô integral*, of f and is denoted by $\int_\alpha^t f(s)dB_s$ $(\alpha \le t \le \beta)$. It will be assumed that $t \to \int_\alpha^t f(s)dB_s$ is continuous for all $\omega \in A$, with an arbitrary specification as a

nonrandom constant on A^c. If $f \in \mathcal{M}[\alpha, \infty)$, then such a continuous version exists for $\int_\alpha^t f(s)dB_s$ on the infinite interval $[\alpha, \infty)$.

It should be noted that the Itô integral of f is well defined up to a null set. For if $\{f_n\}$, $\{g_n\}$ are two sequences of nonanticipative step functionals both satisfying (6.35), then

$$\mathbb{E}\int_\alpha^\beta (f_n(s) - g_n(s))^2 ds \longrightarrow 0. \tag{6.45}$$

If $\{f_{n_k}\}_{k=1}^\infty$, $\{g_{m_k}\}_{k=1}^\infty$ are subsequences of $\{f_n\}_{n=1}^\infty$, $\{g_m\}_{m=1}^\infty$ such that $\int_\alpha^t f_{n_k}(s)dB_s$ and $\int_\alpha^t g_{m_k}(s)dB_s$ both converge uniformly to some processes $Y = \{Y_t : \alpha \le t \le \beta\}$, $Z = \{Z_t : \alpha \le t \le \beta\}$, respectively, outside a set of zero probability, then

$$\mathbb{E}\int_\alpha^\beta (Y_s - Z_s)^2 ds \le (\beta - \alpha) \varliminf_{k\to\infty} \mathbb{E}\int_\alpha^\beta (f_{n_k}(s) - g_{m_k}(s))^2 ds = 0,$$

by (6.45) and Fatou's lemma. As $\{Y_s : \alpha \le s \le \beta\}$, $\{Z_s : \alpha \le s \le \beta\}$ are a.s. continuous, one must have $Y_s = Z_s$ for $\alpha \le s \le \beta$, outside a P-null set.

Note also that $\int_\alpha^t f(s)dB_s$ is $\mathcal{F}_t$-measurable for $\alpha \le t \le \beta$, $f \in \mathcal{M}[\alpha, \beta]$. The following generalization of Proposition 6.2 is almost immediate.

Theorem 6.5 *Properties (a)–(f) of Proposition 6.2 hold for the extension of the Itô integral to $\mathcal{M}[\alpha, \beta]$.*

Proof This follows from the corresponding properties of the approximating step functionals $\{f_{n_k}\}_{k=1}^\infty$, on taking limits (Exercise 4). ■

Example 5 Let $f(s) = B_s$. Then, $f \in \mathcal{M}[\alpha, \infty)$. Fix $t > 0$. Define $f_n(t) = B_t$ and

$$f_n(s) := B_{j2^{-n}t} \qquad \text{if } j2^{-n}t \le s < (j+1)2^{-n}t \quad (0 \le j \le 2^n - 1).$$

Then, writing $B_{j,n} := B_{j2^{-n}t}$,

$$\begin{aligned} \int_0^t f_n(s)dB_s &= \sum_{j=0}^{2^n-1} B_{j,n}(B_{j+1,n} - B_{j,n}) \\ &= -\frac{1}{2}\sum_{j=0}^{2^n-1} (B_{j+1,n} - B_{j,n})^2 - \tfrac{1}{2}B_0^2 + \tfrac{1}{2}B_t^2 \\ &\to -\tfrac{1}{2}t - \tfrac{1}{2}B_0^2 + \tfrac{1}{2}B_t^2 \quad \text{a.s.} \quad (\text{as } n \to \infty). \end{aligned} \tag{6.46}$$

The last convergence follows from Exercise 1. Hence,

$$\int_0^t B_s dB_s = \frac{1}{2}(B_t^2 - B_0^2) - \frac{1}{2}t. \tag{6.47}$$

Notice that this example explicitly demonstrates (i) the almost sure (sample pathwise) convergence of the partial sums defining the integral, via the strong law of large numbers, and (ii) the a.s. uniformity, via the scaling property of the displacements. Theorem 6.4 shows more generally for any $f \in \mathcal{M}[\alpha, \beta]$, there is always a partition, for which this almost sure uniform convergence happens.

As remarked earlier, the formal application of ordinary calculus would yield $\int_0^t B_s dB_s = \frac{1}{2}(B_t^2 - B_0^2)$. Of course, unlike (6.47), and contrary to the basic martingale property requirement for Itô integrals of Theorem 6.5, the usual rules of calculus do not yield a martingale. An alternative method of evaluation of this integral that exploits the martingale structure of the Itô integral (namely, Itô's lemma) will be given in Chapter 8.

Notice that if one replaced the values of f at the left end points of the subintervals by those at the right end points, then one would get, in place of the first sum in (6.46),

$$\sum_{j=0}^{2n-1} B_{j+1,n}(B_{j+1,n} - B_{j,n}) = \frac{1}{2}\sum_{j=0}^{2^n-1}(B_{j+1,n} - B_{j,n})^2 + \frac{1}{2}(B_t^2 - B_0^2),$$

which converges a.s. to $\frac{1}{2}(B_t^2 - B_0^2) + \frac{1}{2}t$. Again, this limit is not a martingale.

The following useful formula is left as Exercise 7. Also see Exercise 8.

Proposition 6.6 (Integration by Parts) *Let $f(t)$ $(\alpha \leq t \leq \beta)$ be a nonrandom continuously differentiable function. Then,*

$$\int_\alpha^\beta f(s)dB_s = f(\beta)B_\beta - f(\alpha)B_\alpha - \int_\alpha^\beta B_s f'(s)ds.$$

Notice that if $f(s)$ is a random nonanticipative functional, then the derivative $f'(s)$ may not be anticipative (assuming it exists).

Exercises

1. (*Quadratic Variation of Brownian Motion*) Let $\{B_t : t \geq 0\}$ be a standard Brownian motion. Prove that, for every $t > 0$, as $n \to \infty$, $\max_{1\leq j\leq 2^n} |\{\sum_{i=1}^{j}(B_{i2^{-n}t} - B_{(i-1)2^{-n}t})^2 - j2^{-n}t\}| \to 0$, *a.s.* [*Hint:* Apply Doob's maximal inequality to the martingale $X_j, 1 \leq j \leq 2^n$, defined within the curly brackets to get $P(\max_{1\leq j\leq 2^n} |X_j| \geq 2^{-\frac{n}{4}}) \leq \mathbb{E}X_{2^n}^2/2^{-\frac{n}{2}} = 2^{-\frac{n}{2}}\mathrm{Var}(\mathrm{B}_1 - \mathrm{B}_0)^2$.]

2. (*Unbounded Variation*) Let B be a standard Brownian motion starting at zero. Define $V_n = \sum_{j=1}^{2^n} |B_{j2^{-n}} - B_{(j-1)2^{-n}}|$.

   (i) Verify that $\mathbb{E}V_n = 2^{\frac{n}{2}}\mathbb{E}|B_1|$. [*Hint:* $2^{\frac{n}{2}}(B_{j2^{-n}} - B_{(j-1)2^{-n}})$ has the same distribution as B_1.]

   (ii) Show that $\mathrm{Var}V_n = \mathrm{Var}|B_1|$.

   (iii) Show that, with probability one, B is of unbounded variation on every interval $[s,t]$, $s < t$. [*Hint:* V_n is a nondecreasing sequence and therefore has a limit. By Chebyshev's inequality, $P(V_n > M) \to 1$ as $n \to \infty$ for any $M > 0$.]

3. Consider the stochastic integral equation (6.8) with $\mu(\cdot)$ Lipschitzian, $|\mu(x) - \mu(y)| \le M|x - y|$. Solve this equation by the method of successive approximations, with the nth approximation $X_t^{(n)}$ given by

$$X_t^{(n)} = x + \int_0^t \mu(X_s^{(n-1)})ds + \sigma(B_t - B_0) \qquad (n \ge 1),\ \ X_t^{(0)} \equiv x.$$

   [*Hint:* For each $t > 0$, write $\delta_n(t) := \max\{|X_s^{(n)} - X_s^{(n-1)}| : 0 \le s \le t\}$. Fix $T > 0$. Then,

$$\delta_n(T) \le M \int_0^T \delta_{n-1}(s)ds \le M^2 \int_0^T ds \int_0^{t_{n-1}} \delta_{n-2}(t_{n-2}))dt_{n-2} \le \cdots$$
$$\le M^{n-1} \int_0^T ds \int_0^{t_{n-1}} dt_{n-2} \cdots \int_0^{t_2} \delta_1(t_1)dt_1 \le M^{n-1}\delta_1(T)T^{n-1}/(n-1)!$$

   Hence, $\sum_n \delta_n(T)$ converges to a finite limit, which implies the uniform convergence of $X^{(n)} = \{X_s^{(n)} : 0 \le s \le T\}$ to a finite and continuous limit $X = \{X_s : 0 \le s \le T\}$.]

4. Prove Theorem 6.5 in detail.
5. Suppose $f_n, f \in \mathcal{M}[\alpha,\beta]$ and $\mathbb{E}\int_\alpha^\beta (f_n(s) - f(s))^2 ds \to 0$ as $n \to \infty$.

   (i) Prove that

$$U_n := \max\{|\int_\alpha^t f_n(s)dB_s - \int_\alpha^t f(s)dB_s| : \alpha \le t \le \beta\} \longrightarrow 0$$

   in probability.

   (ii) If $\{Y(t) : \alpha \le t \le \beta\}$ is a stochastic process with continuous sample paths and $\int_\alpha^t f_n(s)dB_s \to Y(t)$ in probability for all $t \in [\alpha,\beta]$, then prove that

$$P(Y(t) = \int_\alpha^t f(s)dB_s,\ \forall\ t \in [\alpha,\beta]) = 1.$$

6. Prove the convergence in (6.46). [*Hint:* Apply space-time scaling to express the sum as an average of the form $\frac{t}{2^n}\sum_{j=1}^{2^n} Z_{n,j}^2$, where $Z_{n,1},\ldots,Z_{n,2^n}$ are i.i.d. standard normal. Use Borel-Cantelli I to prove a strong law of large numbers for such an average.]
7. Prove the integration by parts formula in Proposition 6.6. [*Hint:* Rewrite the indicated Riemann-type sum in terms of differences of f.]
8. Apply the integration by parts formula to express Lévy's fractional Brownian motion[2] with parameter $h > 0$, $B_t^{(h)} = \int_0^t s^h dB_s$, $t \geq 0$, as a Riemann integral of Brownian motion paths.
9. Use the integration by parts formula and the martingale property of stochastic integrals to show that $tB_t - \int_0^t B_s ds$, $t \geq 0$, is a martingale. Also establish the martingale property directly using properties of conditional expectation.

[2] Lévy (1953) introduced fractional Brownian motion in a general framework of Gaussian processes. His definition was subsequently modified by Mandelbrot and Van Ness (1968) for an application to the Hurst problem in hydrology. See Bhattacharya and Waymire (2022) for a rather comprehensive treatment of this problem.

Chapter 7
Construction of Diffusions as Solutions of Stochastic Differential Equations

In this chapter it is shown how to use Itô's stochastic calculus to specify a diffusion in terms of given drift and dispersion coefficients as a solution to the corresponding stochastic differential equation. This presents a powerful probabilistic alternative to the analytic theory of determining the transition probabilities from semigroups generated by a corresponding infinitesimal operator. The two approaches combine to provide a powerful overall tool of both probabilistic and analytic importance.

In the last chapter a precise meaning was given to the stochastic differential equation (6.3) in terms of its integrated formulation (6.7). The present chapter is devoted to the solution of (6.3) (or (6.7)). The chapter is organized into subsections in which the scope of the theory is gradually extended to broader classes of coefficients and/or higher dimensions.

Specifically, the goals are to (i) establish existence and uniqueness of a continuous process $X = \{X_t : t \geq \alpha\}$ in $\mathcal{M}[\alpha, \infty)$ such that

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t, \quad t > \alpha, X(\alpha) = X_\alpha, \tag{7.1}$$

for given Lipschitz coefficients μ, σ^2 and $\mathcal{F}_\alpha$-measurable random variable X_α in one dimension; (ii) derive the Markov property (from uniqueness) and strong Markov properties (from Markov property and Feller property); (iii) extend to higher dimensional equations of the (suitably interpreted) form

$$d\mathbf{X}_t = \boldsymbol{\mu}(\mathbf{X}_t)dt + \boldsymbol{\sigma}(\mathbf{X}_t)d\mathbf{B}_t, \quad t > \alpha, \mathbf{X}(\alpha) = \mathbf{X}_\alpha, \tag{7.2}$$

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_7

where $\mathbf{X}$ and $\boldsymbol{\mu}$ are column vectors, $\boldsymbol{\sigma}$ is a matrix, and $\mathbf{B}$ a multidimensional standard Brownian motion of consistent dimensions for the indicated algebraic operations; (iv) relax the Lipschitz conditions to *local* Lipschitz conditions for an existence theorem of a *minimal process*, together with the possibility of *explosion*; (v) derive the Markov and strong Markov properties for the minimal process; and (vi) extend (i)–(iv) to time-dependent coefficients $\boldsymbol{\mu}(\mathbf{x}, t)$, $\boldsymbol{\sigma}(\mathbf{x}, t)$.

7.1 Construction of One-Dimensional Diffusions

Let $\mu(x)$ and $\sigma(x)$ be two real–valued functions on $\mathbb{R}$ that are Lipschitzian, i.e., there exists a constant $M > 0$ such that

$$|\mu(x) - \mu(y)| \leq M|x - y|, \qquad |\sigma(x) - \sigma(y)| \leq M|x - y|, \qquad \forall\ x, y. \tag{7.3}$$

The first order of business is to show that Equation (7.1) is valid as a *stochastic integral equation* in the sense of Itô, i.e.,

$$X_t = X_\alpha + \int_\alpha^t \mu(X_s)ds + \int_\alpha^t \sigma(X_s)dB_s, \qquad (t \geq \alpha). \tag{7.4}$$

Theorem 7.1 *Suppose $\mu(\cdot)$, $\sigma(\cdot)$ satisfy (7.3). Then, for each $\mathcal{F}_\alpha$–measurable square integrable random variable X_α, there exists a unique (except on a set of zero probability) continuous nonanticipative functional $\{X_t : t \geq \alpha\}$ belonging to $\mathcal{M}[\alpha, \infty)$ that satisfies (7.4).*

Proof We prove "existence" by the method of successive approximation. Fix $T > \alpha$. Let

$$X_t^{(0)} := X_\alpha, \qquad \alpha \leq t \leq T. \tag{7.5}$$

Define, recursively,

$$X_t^{(n+1)} := X_\alpha + \int_\alpha^t \mu(X_s^{(n)})ds + \int_\alpha^t \sigma(X_s^{(n)})dB_s, \qquad \alpha \leq t \leq T. \tag{7.6}$$

For example,

$$\begin{aligned} X_t^{(1)} &= X_\alpha + \mu(X_\alpha)(t-\alpha) + \sigma(X_\alpha)(B_t - B_\alpha), \\ X_t^{(2)} &= X_\alpha + \int_\alpha^t \mu([X_\alpha + \mu(X_\alpha)(s-\alpha) + \sigma(X_\alpha)(B_s - B_\alpha)])ds \\ &\quad + \int_\alpha^t \sigma([X_\alpha + \mu(X_\alpha)(s-\alpha) + \sigma(X_\alpha)(B_s - B_\alpha)])dB_s,\ \alpha \leq t \leq T. \end{aligned} \tag{7.7}$$

Note that, for each n, $\{X_t^{(n)} : \alpha \le t \le T\}$ is a continuous nonanticipative functional on $[\alpha, T]$. Also, for $n \ge 1$,

$$
\begin{aligned}
&\big(X_t^{(n+1)} - X_t^{(n)}\big)^2 \\
&= [\int_\alpha^t \big(\mu\big(X_s^{(n)}\big) - \mu\big(X_s^{(n-1)}\big)\big)ds + \int_\alpha^t \big(\sigma\big(X_s^{(n)}\big) - \sigma\big(X_s^{(n-1)}\big)\big)dB_s]^2 \\
&\le 2\Big(\int_\alpha^t \big(\mu(X_s^{(n)}\big) - \mu\big(X_s^{(n-1)}\big)\big)ds\Big)^2 + 2\Big(\int_\alpha^t \big(\sigma\big(X_s^{(n)}\big) - \sigma\big(X_s^{(n-1)}\big)\big)dB_s\Big)^2. \quad (7.8)
\end{aligned}
$$

Write

$$D_t^{(n)} := \mathbb{E}(\max_{\alpha \le s \le t} (X_s^{(n)} - X_s^{(n-1)})^2). \tag{7.9}$$

Taking expectations of the maximum in (7.8), over $\alpha \le t \le T$, and using (7.3) we get

$$
\begin{aligned}
D_T^{(n+1)} \le\ & 2\mathbb{E}M^2\Big(\int_\alpha^T |X_s^{(n)} - X_s^{(n-1)}|ds\Big)^2 \\
&+2\mathbb{E} \max_{\alpha \le t \le T} \Big(\int_\alpha^t (\sigma(X_s^{(n)}) - \sigma(X_s^{(n-1)}))dB_s\Big)^2. \qquad (7.10)
\end{aligned}
$$

Now, using the Cauchy–Schwarz inequality, one has

$$
\begin{aligned}
\mathbb{E}\Big(\int_\alpha^T |X_s^{(n)} - X_s^{(n-1)}|ds\Big)^2 &\le (T-\alpha)\mathbb{E}\int_\alpha^T (X_s^{(n)} - X_s^{(n-1)})^2 ds \\
&\le (T-\alpha)\int_\alpha^T D_s^{(n)} ds. \qquad (7.11)
\end{aligned}
$$

Also, by Doob's maximal inequality Theorem 1.14, and Theorem 6.5 (Itô isometry),

$$
\begin{aligned}
&\mathbb{E} \max_{\alpha \le t \le T} \Big(\int_\alpha^t (\sigma(X_s^{(n)}) - \sigma(X_s^{(n-1)}))dB_s\Big)^2 \\
&\quad \le 4\mathbb{E}\Big(\int_\alpha^T (\sigma(X_s^{(n)}) - \sigma(X_s^{(n-1)}))dB_s\Big)^2 = 4\int_\alpha^T \mathbb{E}(\sigma(X_s^{(n)}) - \sigma(X_s^{(n-1)}))^2 ds \\
&\quad \le 4M^2 \int_\alpha^T \mathbb{E}(X_s^{(n)} - X_s^{(n-1)})^2 ds \le 4M^2 \int_\alpha^T D_s^{(n)} ds. \qquad (7.12)
\end{aligned}
$$

Using (7.11) and (7.12) in (7.10), obtain

$$D_T^{(n+1)} \le (2M^2(T-\alpha) + 8M^2)\int_\alpha^T D_s^{(n)} ds = c_1 \int_\alpha^T D_s^{(n)} ds \qquad (n \ge 1), \tag{7.13}$$

say. An analogous, but simpler, calculation gives

$$D_T^{(1)} \le 2(T-\alpha)^2\mathbb{E}\mu^2(X_\alpha) + 8(T-\alpha)\mathbb{E}\sigma^2(X_\alpha) = c_2, \tag{7.14}$$

where c_2 is a finite positive number (Exercise 5). It follows from (7.13), (7.14), and induction that for $n = 0, 1, \ldots,$

$$D_T^{(n+1)} \le c_1^n \int_\alpha^T ds_n \int_\alpha^{s_n} ds_{n-1} \cdots \int_\alpha^{s_2} D_{s_1}^{(1)} ds_2 \le c_2 \frac{(T-\alpha)^n c_1^n}{n!}. \tag{7.15}$$

By Chebyshev's inequality,

$$P(\max_{\alpha \le t \le T} |X_t^{(n+1)} - X_t^{(n)}| > 2^{-n}) \le \frac{2^{2n} c_2 (T-\alpha)^n c_1^n}{n!}. \tag{7.16}$$

Since the right side is summable in n, it follows from the Borel–Cantelli lemma that

$$P(\max_{\alpha \le t \le T} |X_t^{(n+1)} - X_t^{(n)}| > 2^{-n} \text{ for infinitely many } n) = 0. \tag{7.17}$$

Let N denote the set within parentheses in (7.17). Outside N, the series

$$X_t^{(0)} + \sum_{n=0}^{\infty} (X_t^{(n+1)} - X_t^{(n)}) = \lim_{n\to\infty} X_t^{(n)} \tag{7.18}$$

converges absolutely, uniformly for $\alpha \le t \le T$, to a nonanticipative functional $\{X_t : \alpha \le t \le T\}$, say. Since each $X_t^{(n)}$ is continuous and the convergence is uniform, the limit is continuous.

For continuous nonanticipative functionals f, g on $[\alpha, T]$, define the norm (when finite):

$$||f - g||^2 := \mathbb{E} \max_{\alpha \le s \le T} (f(s) - g(s))^2. \tag{7.19}$$

Writing $X = \{X_s : \alpha \le s \le T\}$, $X^{(n)} = \{X_s^{(n)} : \alpha \le s \le T\}$, one has (see (7.15), (7.18)),

$$||X - X^{(n)}|| \le \sum_{m=n}^{\infty} ||X^{(m+1)} - X^{(m)}|| \le (c_2 \sum_{m=n}^{\infty} \frac{(T-\alpha)^m c_1^m}{m!})^{\frac{1}{2}} \to 0. \tag{7.20}$$

In particular, $||X|| < \infty$, and $X \in \mathcal{M}[\alpha, T]$.

To prove that X satisfies (7.4), note that

$$\begin{aligned}
&\max_{\alpha \le t \le T} \left| \int_\alpha^t \mu(X_s^{(n)})ds + \int_\alpha^t \sigma(X_s^{(n)})dB_s - \int_\alpha^t \mu(X_s)ds - \int_\alpha^t \sigma(X_s)dB_s \right| \\
&\le M \int_\alpha^T |X_s^{(n)} - X_s|ds + \max_{\alpha \le t \le T} \left| \int_\alpha^t (\sigma(X_s^{(n)}) - \sigma(X_s))dB_s \right| \\
&\le M(T-\alpha) \max_{\alpha \le s \le T} |X_s^{(n)} - X_s| \\
&\quad + \max_{\alpha \le t \le T} \left| \int_\alpha^t (\sigma(X_s^{(n)}) - \sigma(X_s))dB_s \right|. \qquad (7.21)
\end{aligned}$$

The first term on the right side goes to zero by the a.s. uniform convergence proved above. The second term goes to zero a.s. as $n \to \infty$ by Theorem 6.4 and (7.20). Thus the right side of (7.6) converges uniformly to the right side of (7.4) as $n \to \infty$. Since $X^{(n)}$ converges uniformly to X on $[\alpha, T]$ outside a set of zero probability, (7.4) holds on $[\alpha, T]$ with probability one.

It remains to prove the *uniqueness* to the solution to (7.4). Let $Y = \{Y_t : \alpha \le t \le T\}$ be another solution. Then write

$$\varphi_t := \mathbb{E}(\max\{|X_s - Y_s| : \alpha \le s \le t\})^2$$

and use (7.9)–(7.13) with φ_t in place of $D_t^{(n)}$, X_t in place of $X_t^{(n)}$ and Y_t in place of $X_t^{(n-1)}$, to get

$$\varphi_T \le c_1 \int_\alpha^T \varphi_s ds. \qquad (7.22)$$

Since $t \to \varphi_t$ is nondecreasing, iteration of (7.22) leads, just as in (7.14), (7.15), to

$$\varphi_T \le \frac{c_2(T-\alpha)^n c_1^n}{n!} \longrightarrow 0 \quad \text{as } n \to \infty. \qquad (7.23)$$

Hence, $\varphi_T = 0$, i.e., $X_t = Y_t$ on $[\alpha, T]$ outside a set of zero probability. ■

The next result identifies the stochastic process $X = \{X_t : t \ge 0\}$ solving (7.4), in the case $\alpha = 0$, as a diffusion with drift $\mu(\cdot)$, diffusion coefficient $\sigma^2(\cdot)$, and initial distribution as the distribution of the given random variable X_0.

Theorem 7.2 *Assume (7.3). For each $x \in \mathbb{R}$ let $X^x = \{X_t^x : t \ge 0\}$ denote the unique continuous solution in $\mathcal{M}[0, \infty)$ of Itô's integral equation*

$$X_t^x = x + \int_0^t \mu(X_s^x)ds + \int_0^t \sigma(X_s^x)dB_s \qquad (t \ge 0). \qquad (7.24)$$

Then $X^x = \{X_t^x : t \ge 0\}$ is a Markov process.

Proof By the additivity of the Riemann and stochastic integrals,

$$X_t^x = X_s^x + \int_s^t \mu(X_u^x)du + \int_s^t \sigma(X_u^x)dB_u, \qquad (t \geq s). \tag{7.25}$$

Consider also the equation, for $z \in \mathbb{R}$,

$$X_t = z + \int_s^t \mu(X_u)du + \int_s^t \sigma(X_u)dB_u, \qquad (t \geq s). \tag{7.26}$$

Let us write the solution to (7.26) (in $\mathcal{M}[s,\infty)$) as a.s. equal to a function $\theta(s,t;z,B_s^t)$ where $B_s^t := \{B_u - B_s : s \leq u \leq t\}$. It may be seen from the successive approximation scheme (7.5)–(7.7) that $\theta(s,t;z,B_s^t)$ is *measurable* in (z, B_s^t) with respect to $\mathcal{B}(\mathbb{R}) \otimes \sigma(B_s^t)$ (Exercise 6). As $\{X_u^x : u \geq s\}$ is a continuous stochastic process in $\mathcal{M}[s,\infty)$ and is, by (7.25), a solution to (7.26) with $z = X_s^x$, it follows from the *uniqueness* of this solution that

$$X_t^x = \theta(s,t;X_s^x, B_s^t), \qquad (t \geq s), \text{ a.s.} \tag{7.27}$$

Since X_s^x is $\mathcal{F}_s$–measurable and $\mathcal{F}_s$ and B_s^t are independent, (7.27) implies, using the substitution property for conditional expectation,[1] that the conditional distribution of X_t^x given $\mathcal{F}_s$ is the distribution of the random variable $\theta(s,t;z,B_s^t)$, say $q(s,t;z,d\theta)$, evaluated at $z = X_s^x$. Since $\sigma\{X_u^x : 0 \leq u \leq s\} \subseteq \mathcal{F}_s$, the Markov property is proved. To prove *homogeneity* of this Markov process, consider for every $h > 0$, the solution $\theta(s+h, t+h; z, B_{s+h}^{t+h})$ of

$$X_{t+h} = z + \int_{s+h}^{t+h} \mu(X_u)du + \int_{s+h}^{t+h} \sigma(X_u)dB_u, \qquad (t \geq s). \tag{7.28}$$

Writing $Y_u = X_{u+h}$ and with the change of variables $u' = u - h$, (7.28) may be expressed as

$$\begin{aligned} Y_t &= z + \int_s^t \mu(X_{u'+h})du' + \int_s^t \sigma(X_{u'+h})dB_{u'+h} \\ &= z + \int_s^t \mu(Y_{u'})du' + \int_s^t \sigma(Y_{u'})d\tilde{B}_{u'}, \qquad (t \geq s), \end{aligned} \tag{7.29}$$

where $\tilde{B}_{u'} = B_{u'+h}$, so that the process of the increments $\{\tilde{B}_{u'} - \tilde{B}_s \equiv B_{u'+h} - B_{s+h} : s \leq u' \leq t\} \equiv \tilde{B}_s^t$ has the same distribution as B_s^t. Hence Y_t has the

[1] See BCPT,[2] p. 38.

[2] Throughout, BCPT refers to Bhattacharya and Waymire (2016), A Basic Course in Probability Theory.

same distribution as X_t, i.e., $q(s+h, t+h; z, d\theta) = q(s, t; z, d\theta)$. This proves homogeneity. ∎

Definition 7.1 The family of Markov processes $X^z = \{X_t^z : t \geq 0\}, z \in \mathbb{R}$, solving (7.4) (with $\alpha = 0$) for respective initial values $X_0 = z \in \mathbb{R}$, defines a *diffusion on* $\mathbb{R}$ with drift $\mu(\cdot)$ and diffusion parameter (coefficient) $\sigma(\cdot)$, starting at $z \in \mathbb{R}$.

A Word on Terminology To be consistent with the predominant use in the literature, some abuse of terminology in reference to $\sigma(\cdot)$ vs $\sigma^2(\cdot)$ as the 'diffusion coefficient' will occur in the text. This is largely due to the explicit appearance of $\sigma(\cdot)$ as a 'coefficient' in the stochastic differential equation and $\sigma^2(\cdot)$ as a 'coefficient' in the associated infinitesimal generator (see (7.30) below). As a result, some reliance on the context becomes necessary; however, we shall refer to $\sigma(\cdot)$ as the diffusion parameter in instances where it may not be clear otherwise.

We will denote the transition probability of the diffusions $X^z = \{X_t^z : t \geq 0\}$, $z \in \mathbb{R}$ by $p(t; x, dy)$, $(x \in \mathbb{R}, t \geq 0)$.

Under the assumption (7.3) and in keeping with the "locally Brownian" intuitive description of the diffusion, one may compute the infinitesimal moments in terms of the drift and diffusion coefficients as follows (Exercise 8):

$$(i)\ \mathbb{E}(X_t^x - x) = \mu(x)t + o(t),\ t \downarrow 0. \tag{7.30}$$

$$(ii)\ \mathrm{Var}(X_t^x) = \mathbb{E}(X_t^x - x)^2 + o(t) = \sigma^2(x)t + o(t),\ t \downarrow 0. \tag{7.31}$$

In the next chapter (using Itô's lemma), it will also follow that

$$P(|X_t^x - x| > \varepsilon) = o(t),\ t \downarrow 0. \tag{7.32}$$

Example 1 (Ornstein–Uhlenbeck Process and the Langevin Equation) Let $\mu(x) = -\gamma x$, $\sigma(x) = \sigma$, $x \in \mathbb{R}$. Then the successive approximations $X_t^{(n)}$ (see Equations (7.5)–(7.7)) are given by (assuming $B_0 = 0$)

$$\begin{aligned}
X_t^{(0)} &\equiv X_0,\\
X_t^{(1)} &= X_0 + \int_0^t -\gamma X_0 ds + \sigma B_t = X_0(1 - t\gamma) + \sigma B_t,\\
X_t^{(2)} &= X_0 + \int_0^t -\gamma\{X_0(1 - s\gamma) + \sigma B_s\}ds + \sigma B_t\\
&= (1 - t\gamma + \frac{t^2\gamma^2}{2!})X_0 - \gamma\sigma \int_0^t B_s ds + \sigma B_t,
\end{aligned}$$

$$
\begin{aligned}
X_t^{(3)} &= X_0 + \int_0^t -\gamma\{(1 - s\gamma + \frac{s^2\gamma^2}{2!})X_0 - \gamma\sigma \int_0^s B_u du + \sigma B_s\} ds + \sigma B_t \\
&= (1 - t\gamma + \frac{t^2\gamma^2}{2!} - \frac{t^3\gamma^3}{3!})X_0 + \gamma^2\sigma \int_0^t \int_0^s B_u du\, ds - \gamma\sigma \int_0^t B_s ds + \sigma B_t \\
&= (1 - t\gamma + \frac{t^2\gamma^2}{2!} - \frac{t^3\gamma^3}{3!})X_0 + \gamma^2\sigma \int_0^t (\int_u^t ds) B_u du - \gamma\sigma \int_0^t B_s ds + \sigma B_t \\
&= (1 - t\gamma + \frac{t^2\gamma^2}{2!} - \frac{t^3\gamma^3}{3!})X_0 + \gamma^2\sigma \int_0^t (t-u) B_u du - \gamma\sigma \int_0^t B_s ds + \sigma B_t.
\end{aligned}
\tag{7.33}
$$

Assume, as an induction hypothesis,

$$
X_t^{(n)} = (\sum_{m=0}^{n} \frac{(-t\gamma)^m}{m!})X_0 + \sum_{m=1}^{n-1} (-\gamma)^m \sigma \int_0^t \frac{(t-s)^{m-1}}{(m-1)!} B_s ds + \sigma B_t. \tag{7.34}
$$

Now use (7.6) to check that (7.34) holds for $n+1$, replacing n. Therefore, (7.34) holds for all n. But as $n \to \infty$, the right side converges (for every $\omega \in \Omega$) to

$$
e^{-t\gamma} X_0 - \gamma\sigma \int_0^t e^{-\gamma(t-s)} B_s ds + \sigma B_t. \tag{7.35}
$$

Hence, X_t equals (7.35). In particular, with $X_0 = x$,

$$
X_t^x = e^{-t\gamma} x - \gamma\sigma \int_0^t e^{-\gamma(t-s)} B_s ds + \sigma B_t. \tag{7.36}
$$

As a special case, for $\gamma > 0$, $\sigma \neq 0$, (7.36) gives a representation of an Ornstein–Uhlenbeck process as a functional of a Brownian motion. The corresponding stochastic differential equation for this choice of drift and diffusion parameters is referred to as the *Langevin equation*.

7.2 Extension to Multidimensional Diffusions

In order to construct multidimensional diffusions by the method of Itô it is necessary to define stochastic integrals for vector–valued integrands with respect to the increments of a multidimensional Brownian motion. This turns out to be rather straightforward.

Let $\mathbf{B} = \{\mathbf{B}_t = (B_t^{(1)}, \ldots, B_t^{(k)}) : t \geq 0\}$ be a standard k-dimensional Brownian motion with respect to a filtration $\{\mathcal{F}_t : t \geq 0\}$. For example, one may take $\mathcal{F}_t = \sigma(\mathbf{B}_s : s \leq t)$. Assume (6.9) *for* $\{\mathbf{B}_t : t \geq 0\}$. Define a vector–valued stochastic process

$$\mathbf{f} = \{\mathbf{f}(t) = (f^{(1)}(t), \ldots, f^{(k)}(t)) : \alpha \le t \le \beta\}$$

to be a *nonanticipative step functional on* $[\alpha, \beta]$ if (6.12) holds for some finite set of time points $t_0 = \alpha < t_1 < \cdots < t_m = \beta$ and some $\mathcal{F}_{t_i}$-measurable random vectors $\mathbf{f}_i$ $(0 \le i \le m)$ with values in $\mathbb{R}^k$.

Definition 7.2 (Stochastic Integral) The *stochastic integral,* or the *Itô integral* of a nonanticipative step functional $\mathbf{f}$ on $[\alpha, \beta]$ is defined by

$$\int_\alpha^t \mathbf{f}(s) \cdot d\mathbf{B}_s$$

$$:= \begin{cases} \mathbf{f}(t_0) \cdot (\mathbf{B}_t - \mathbf{B}_{t_0}) \text{ for } t \in [\alpha, t_1], \\ \sum_{j=1}^{i} \mathbf{f}(t_{j-1}) \cdot (\mathbf{B}_{t_j} - \mathbf{B}_{t_{j-1}}) + \mathbf{f}(t_i) \cdot (\mathbf{B}_t - \mathbf{B}_{t_i}) \text{ for } t \in (t_i, t_{i+1}]. \end{cases} \tag{7.37}$$

Here $\cdot$ (*dot*) denotes the Euclidean inner product,

$$\mathbf{f}(s) \cdot (\mathbf{B}_t - \mathbf{B}_s) = \sum_{i=1}^{k} f^{(i)}(s)(B_t^{(i)} - B_s^{(i)}), \tag{7.38}$$

with $\mathbf{f}(s) = (f^{(1)}(s), \ldots, f^{(k)}(s))$.

It follows from this definition that the stochastic integral (7.37) is the sum of k one-dimensional stochastic integrals,

$$\int_\alpha^t \mathbf{f}(s) \cdot d\mathbf{B}_s = \sum_{i=1}^{k} \int_\alpha^t f^{(i)}(s) dB_s^{(i)}. \tag{7.39}$$

Parts (a), (b) of Proposition 6.2 extend immediately. In order to extend part (c) assume, as in (6.16),

$$\int_\alpha^\beta \mathbb{E}|\mathbf{f}(u)|^2 du \equiv \sum_{i=1}^{k} \int_\alpha^\beta \mathbb{E}(f^{(i)}(u))^2 du < \infty. \tag{7.40}$$

The square integrability and the martingale property of the stochastic integral follow from those of each of its k summands in (7.39). Similarly, one has

$$\mathbb{E}(\int_\alpha^t \mathbf{f}(s) \cdot d\mathbf{B}_s | \mathcal{F}_\alpha) = 0, \ \mathbb{E}(\int_\alpha^t \mathbf{f}(s) \cdot d\mathbf{B}_s) = 0. \tag{7.41}$$

It remains to prove the analog of (6.19), namely,

$$\mathbb{E}((\int_\alpha^t \mathbf{f}(u) \cdot d\mathbf{B}_u)^2 | \mathcal{F}_\alpha) = \mathbb{E}(\int_\alpha^t |\mathbf{f}(u)|^2 du | \mathcal{F}_\alpha). \tag{7.42}$$

In view of (6.19) applied to $\int_\alpha^t f^{(i)}(u)dB_u^{(i)}$ $(1 \le i \le k)$, it is enough to prove that the product terms vanish, i.e.,

$$\mathbb{E}[(\int_\alpha^t f^{(i)}(u)dB_u^{(i)})(\int_\alpha^t f^{(j)}(u)dB_u^{(j)})|\mathcal{F}_\alpha] = 0 \quad \text{for } i \ne j. \tag{7.43}$$

To see this, note that, for $s < s' < u < u'$,

$$\begin{aligned}
&\mathbb{E}[f^{(i)}(s)(B_{s'}^{(i)} - B_s^{(i)})f^{(j)}(u)(B_{u'}^{(j)} - B_u^{(j)}) \mid \mathcal{F}_\alpha] \\
&\quad = \mathbb{E}[\mathbb{E}(f^{(i)}(s)(B_{s'}^{(i)} - B_s^{(i)})f^{(j)}(u)(B_{u'}^{(j)} - B_u^{(j)}) \mid \mathcal{F}_u) \mid \mathcal{F}_\alpha] \\
&\quad = \mathbb{E}[f^{(i)}(s)(B_{s'}^{(i)} - B_s^{(i)})f^{(j)}(u)\mathbb{E}(B_{u'}^{(j)} - B_u^{(j)} \mid \mathcal{F}_u) \mid \mathcal{F}_\alpha] = 0.
\end{aligned} \tag{7.44}$$

Also, by the independence of $\mathcal{F}_s$ and $\{\mathbf{B}_{s'} - \mathbf{B}_s : s' \ge s\}$, and by the independence of $\{B_t^{(i)} : t \ge 0\}$ and $\{B_t^{(j)} : t \ge 0\}$ for $i \ne j$,

$$\begin{aligned}
&\mathbb{E}[f^{(i)}(s)(B_{s'}^{(i)} - B_s^{(i)})f^{(j)}(s)(B_{s'}^{(j)} - B_s^{(j)}) \mid \mathcal{F}_\alpha] \\
&\quad = \mathbb{E}[\mathbb{E}(f^{(i)}(s)(B_{s'}^{(i)} - B_s^{(i)})f^{(j)}(s)(B_{s'}^{(j)} - B_s^{(j)}) \mid \mathcal{F}_s) \mid \mathcal{F}_\alpha] \\
&\quad = \mathbb{E}[f^{(i)}(s)f^{(j)}(s)\{\mathbb{E}((B_{s'}^{(i)} - B_s^{(i)})(B_{s'}^{(j)} - B_s^{(j)}) \mid \mathcal{F}_s)\} \mid \mathcal{F}_\alpha] = 0.
\end{aligned}$$

One may similarly prove, for all $f, g \in \mathcal{M}[\alpha, T]$, $\alpha \le t \le T$,

$$\mathbb{E}((\int_\alpha^t f(s)dB_s)(\int_\alpha^t g(s)d\,B_s)|\mathcal{F}_\alpha) = \mathbb{E}(\int_\alpha^t f(s)g(s)ds \mid \mathcal{F}_\alpha). \tag{7.45}$$

Thus, Proposition 6.2 is fully extended to k-dimension. For the extension to more general nonanticipative functionals, consider the following analog of Definition 6.2

Definition 7.3 If $\mathbf{f} = \{\mathbf{f}(t) = (f^{(1)}(t), \dots, f^{(k)}(t)) : \alpha \le t \le \beta\}$ is a progressively measurable stochastic process with values in $\mathbb{R}^k$ such that $\mathbf{f}(t)$ is $\mathcal{F}_t$–measurable for all $t \in [\alpha, \beta]$, then $\mathbf{f}$ is called a *nonanticipative functional on* $[\alpha, \beta]$ *with values in* $\mathbb{R}^k$. If, in addition,

$$\mathbb{E}\int_\alpha^\beta |\mathbf{f}(t)|^2 dt < \infty, \tag{7.46}$$

then $\mathbf{f}$ is said to *belong to* $\mathcal{M}[\alpha, \beta]$. If $\mathbf{f} \in \mathcal{M}[\alpha, \beta]$ for all $\beta > \alpha$, then $\mathbf{f}$ *belongs to* $\mathcal{M}[\alpha, \infty)$.

For $\mathbf{f} \in \mathcal{M}[\alpha, \beta]$ with values in $\mathbb{R}^k$ one may apply the results in Chapter 6 to each of the k components $\int_\alpha^t f^{(i)}(u)dB_u^{(i)}$ $(1 \le i \le k)$ to extend Proposition 6.3 to k-dimension and to define the stochastic integral $\int_\alpha^t \mathbf{f}(u)d\mathbf{B}_u$ for a k-dimensional

$\mathbf{f} \in \mathcal{M}[\alpha, \beta]$. The following extension of Theorem 6.5 on Itô isometry is now immediate.

Theorem 7.3 *Suppose* $\mathbf{f}, \mathbf{g} \in \mathcal{M}[\alpha, \beta]$ *with values in* $\mathbb{R}^k$. *Then*

a. $t \to \int_\alpha^t \mathbf{f}(u) \cdot d\mathbf{B}_u$ *is continuous and additive.*

b. $\left\{\int_\alpha^t \mathbf{f}(u) \cdot d\mathbf{B}_u : \alpha \le t \le \beta\right\}$ *is a square integrable* $\{\mathcal{F}_t : \alpha \le t \le \beta\}$*–martingale and*

$$\mathbb{E}((\int_\alpha^t \mathbf{f}(u) \cdot d\mathbf{B}_u)^2 | \mathcal{F}_\alpha) = \mathbb{E}(\int_\alpha^t |\mathbf{f}(u)|^2 du | \mathcal{F}_\alpha), \tag{7.47}$$

and $(\int_\alpha^t \mathbf{f}(u) \cdot d\mathbf{B}_u)^2 - \int_\alpha^t |\mathbf{f}(u)|^2 du$, $\alpha \le t \le \beta$, *is a martingale.*

c.

$$\mathbb{E}[\int_\alpha^t \mathbf{f}(s) d\mathbf{B}_s \int_\alpha^t \mathbf{g}(s) d\mathbf{B}_s | \mathcal{F}_\alpha] = \int_\alpha^t \mathbb{E}[\mathbf{f}(s) \cdot \mathbf{g}(s) | \mathcal{F}_\alpha] ds \tag{7.48}$$

d. *(Itô Isometry)* $\mathbb{E}[\int_\alpha^t \mathbf{f}(s) d\mathbf{B}_s \int_\alpha^t \mathbf{g}(s) d\mathbf{B}_s] = \int_\alpha^t \mathbb{E}[\mathbf{f}(s) \cdot \mathbf{g}(s)] ds$ *if* f, g *both satisfy (7.46).*

e. *If* $r \ne r'$, $f, g \in M[\alpha, \beta]$ *are real-valued, then*

$$\mathbb{E}[\int_\alpha^t f(s) dB_s^{(r)} \int_\alpha^t g(s) dB_s^{(r')}] = 0, \ \alpha \le t \le T.$$

The final task is to construct multidimensional diffusions. For this let $\mu^{(i)}(\mathbf{x})$, $\sigma_{ij}(\mathbf{x})$ $(1 \le i, j \le k)$ be real-valued Lipschitzian functions on $\mathbb{R}^k$. Then, writing, $\boldsymbol{\sigma}(\mathbf{x})$ for the $k \times k$ matrix $((\sigma_{ij}(\mathbf{x})))$,

$$|\boldsymbol{\mu}(\mathbf{x}) - \boldsymbol{\mu}(\mathbf{y})| \le M|\mathbf{x} - \mathbf{y}|, \quad \|\boldsymbol{\sigma}(\mathbf{x}) - \boldsymbol{\sigma}(\mathbf{y})\| \le M|\mathbf{x} - \mathbf{y}|, \quad (\mathbf{x}, \mathbf{y} \in \mathbb{R}^k), \tag{7.49}$$

for some positive constant M. Here $\|\cdot\|$ denotes the *matrix norm,*

$$\|\mathbf{D}\| := \sup\{|\mathbf{D}\mathbf{z}| : \mathbf{z} \in \mathbb{R}^k, |\mathbf{z}| = 1\}.$$

Consider the k-dimensional stochastic differential equation

$$d\mathbf{X}_t = \boldsymbol{\mu}(\mathbf{X}_t) dt + \boldsymbol{\sigma}(\mathbf{X}_t) d\mathbf{B}_t, \quad (t \ge \alpha), \tag{7.50}$$

defined by the vector stochastic integral equation,

$$\mathbf{X}_t = \mathbf{X}_\alpha + \int_\alpha^t \boldsymbol{\mu}(\mathbf{X}_s) ds + \int_\alpha^t \boldsymbol{\sigma}(\mathbf{X}_s) d\mathbf{B}_s, \quad (t \ge s), \tag{7.51}$$

which, in turn, is a shorthand for the system of k equations

$$X_t^{(i)} = X_\alpha^{(i)} + \int_\alpha^t \boldsymbol{\mu}^{(i)}(\mathbf{X}_s)ds + \int_\alpha^t \boldsymbol{\sigma}_i(\mathbf{X}_s) \cdot d\mathbf{B}_s, \quad (1 \le i \le k). \tag{7.52}$$

Here $\boldsymbol{\sigma}_i(\mathbf{x})$ is the k-dimensional vector, $(\sigma_{i1}(\mathbf{x}), \dots, \sigma_{ik}(\mathbf{x}))$. The proofs of the following theorems are entirely analogous to those of Theorems 7.1 and 7.2 (Exercises 3 and 4). Basically, one only needs to write

$$|\mathbf{X}_t^{(n+1)} - \mathbf{X}_t^{(n)}|^2 \quad \text{instead of} \quad (X_t^{(n+1)} - X_t^{(n)})^2,$$

$$|\int_\alpha^t (\boldsymbol{\mu}(\mathbf{X}_s^{(n)}) - \boldsymbol{\mu}(\mathbf{X}_s^{(n-1)}))ds|^2 \ \text{ in place of } \ (\int_\alpha^t (\mu(X_s^{(n)}) - \mu(X_s^{(n-1)}))ds)^2$$

$$\|\boldsymbol{\sigma}(\mathbf{X}_s^{(n)}) - \boldsymbol{\sigma}(\mathbf{X}_s^{(n-1)})\|^2, \quad \text{for} \quad (\sigma(X_s^{(n)}) - \sigma(X_s^{(n-1)})^2, \quad \text{etc.}$$

Theorem 7.4 *Suppose $\boldsymbol{\mu}(\cdot), \boldsymbol{\sigma}(\cdot)$ satisfy (7.49). Then, for each $\mathcal{F}_\alpha$–measurable square integrable random vector $\mathbf{X}_\alpha$, there exists a unique (up to a P–null set) continuous element $\mathbf{X} = \{\mathbf{X}_t : t \ge \alpha\}$ of $\mathcal{M}[\alpha, \infty)$ such that (7.51) holds.*

Theorem 7.5 *Suppose $\boldsymbol{\mu}(\cdot), \boldsymbol{\sigma}(\cdot)$ satisfy (7.49), and let $\mathbf{X}^{\mathbf{x}} = \{\mathbf{X}_t^{\mathbf{x}} : t \ge 0\}$ denote the unique (up to a P–null set) continuous nonanticipative functional in $\mathcal{M}[0, \infty)$ satisfying the integral equation*

$$\mathbf{X}_t = \mathbf{x} + \int_0^t \boldsymbol{\mu}(\mathbf{X}_s)ds + \int_0^t \boldsymbol{\sigma}(\mathbf{X}_s)d\mathbf{B}_s, \quad (t \ge 0). \tag{7.53}$$

Then $\mathbf{X}^{\mathbf{x}} = \{\mathbf{X}_t^{\mathbf{x}} : t \ge 0\}$ is a Markov process.

Definition 7.4 The family of Markov processes $\mathbf{X}^{\mathbf{x}} = \{\mathbf{X}_t^{\mathbf{x}} : t \ge 0\}, \mathbf{x} \in \mathbb{R}^k$ define *diffusion on* $\mathbb{R}^k$ with drift $\boldsymbol{\mu}(\cdot)$ and diffusion matrix $\boldsymbol{\sigma}(\cdot)\boldsymbol{\sigma}'(\cdot)$, $\boldsymbol{\sigma}'(\cdot)$ being the transpose $\boldsymbol{\sigma}(\cdot)$.

Remark 7.1 It may be noted that in Theorems 7.4 and 7.5 it is not assumed that $\boldsymbol{\sigma}(\mathbf{x})$ is positive definite. It is known from the theory of partial differential equations that the positive definiteness guarantees the existence of a *density* for the transition probability and its *smoothness* (see Friedman 1964), but it is not needed for the Markov property.

7.3 An Extension of the Itô Integral & SDEs with Locally Lipschitz Coefficients

In this subsection we present a further extension of the previously developed theory to accommodate the construction of diffusions with *locally Lipschitz* drift and diffusion coefficients (parameters): For each n, there is a constant M_n such that for $|\mathbf{x}|, |\mathbf{y}| \le n$,

$$|\boldsymbol{\mu}(\mathbf{x}) - \boldsymbol{\mu}(\mathbf{y})| \leq M_n|\mathbf{x} - \mathbf{y}| \quad \text{and } |\boldsymbol{\sigma}(\mathbf{x}) - \boldsymbol{\sigma}(\mathbf{y})| \leq M_n|\mathbf{x} - \mathbf{y}|. \tag{7.54}$$

Example 2 For a simple example in anticipation of new phenomena, consider $\mu(x) = x^2, x \in \mathbb{R}$, with $\sigma^2 \equiv 0$. One has $|\mu(x)-\mu(y)| = |x+y||x-y| \leq 2n|x-y|$ for $|x| \leq n, |y| \leq n$. Observe that the solution to $dX(t) = X^2(t)dt, X(0) = 1$, namely, the process $X(t) = (1-t)^{-1}, 0 \leq t < 1$, "explodes"out of the state space at $t = 1$.

The following result provides a natural way in which to view integration over intervals determined by a bounded stopping time.

Proposition 7.6 *Let $f \in \mathcal{M}[\alpha, \beta]$ and τ a (bounded) stopping time with values in $[\alpha, \beta]$. Then, writing*

$$I_f(t) := \int_\alpha^t f(s)dB(s), \qquad \alpha \leq t \leq \beta, \tag{7.55}$$

one has

$$I_f(\tau) = \int_\alpha^\beta f(s)\mathbf{1}_{[\tau \geq s]}dB(s). \tag{7.56}$$

Proof First note that $s \to \mathbf{1}_{[\tau \geq s]}$ is left-continuous on $[\alpha, \beta]$ and, therefore, $(s, \omega) \to \mathbf{1}_{[\tau(\omega) \geq s]}$ is nonanticipative, i.e., progressively measurable. Thus $g(s) := f(s)\mathbf{1}_{[\tau \geq s]} \in \mathcal{M}[\alpha, \beta]$. Assume first that τ has finitely many values s_i $(1 \leq i \leq k)$, $s_1 < s_2 < \cdots < s_k$ in $[\alpha, \beta]$. Then

$$\mathbf{1}_{[\tau \geq s]} = \begin{cases} 1 & \text{for } s \in [\alpha, s_1], \\ \mathbf{1}_{[\tau \geq s_{i+1}]} & \text{for } s \in (s_i, s_{i+1})](i = 1, 2, \cdots, k-1), \\ 0 & \text{for } s \in (s_k, \beta)]. \end{cases}$$

By Theorem 7.3(a), writing $g(s) = f(s)\mathbf{1}_{[\tau \geq s]}$, $I_g(t) = \int_\alpha^t g(s)dB(s)$, one has

$$\begin{aligned} I_g(s_1) &= \int_\alpha^{s_1} f(s)dB(s), \\ I_g(s_j) &= \int_\alpha^{s_1} f(s)dB(s) + \sum_{i=1}^{j-1} \int_{s_i}^{s_{i+1}} f(s)\mathbf{1}_{[\tau \geq s_{i+1}]}dB(s) \\ &= \int_\alpha^{s_1} f(s)dB(s) + \sum_{i=1}^{j-1} \mathbf{1}_{[\tau \geq s_{i+1}]} \int_{s_i}^{s_{i+1}} f(s)dB(s), \quad (j = 2, \cdots, k), \\ I_g(\beta) &= I_g(s_k). \end{aligned} \tag{7.57}$$

Since on $[\tau = s_j]$, one has $\mathbf{1}_{[\tau \geq s_{i+1}]} = 1$ for $i = 1, 2, \cdots j-1$, and $\mathbf{1}_{[\tau \geq s_{i+1}]} = 0$ for $i \geq j$,

$$
\begin{aligned}
I_g(\beta) &= \int_\alpha^{s_1} f(s)dB(s) + \sum_{i=1}^{j-1} \int_{s_i}^{s_{i+1}} f(s)dB(s) \\
&= \int_\alpha^{s_j} f(s)dB(s) = I_f(s_j) = I_f(\tau) \quad (j = 1, \ldots, k) \quad \text{on } [\tau = s_j],
\end{aligned}
$$

proving (7.56). For general τ (with values in $[\alpha, \beta]$) define the approximation $\tau^{(n)} = \alpha + (\beta - \alpha)i/2^n$ on the event $[\alpha + (\beta - \alpha)(i-1)/2^n \leq \tau < \alpha + (\beta - \alpha)i/2^n]$ $(i = 1, 2, \cdots, 2^n)$, $\tau^{(n)} = \beta$ on $[\tau = \beta]$. Then $\tau^{(n)} \downarrow \tau$, and $[\tau^{(n)} \geq s] \downarrow [\tau \geq s]$ for $s \in [\alpha, \beta]$, as $n \uparrow \infty$. Writing $g_n(s, \omega) = f(s)\mathbf{1}_{[\tau^{(n)}(\omega) \geq s]}$, one then has

$$
\mathbb{E} \int_\alpha^\beta (g_n(s) - g(s))^2 ds \longrightarrow 0 \qquad \text{as } n \to \infty,
$$

i.e., $g_n \to g$ in $\mathcal{M}[\alpha, \beta]$. Hence, there exists a sequence of integers $n_1 < n_2 < \cdots$ such that $I_{g_{n_k}}(t) \to I_g(t)$ uniformly over $t \in [\alpha, \beta]$ as $k \to \infty$, outside a P–null set. But $I_{g_{n_k}}(\beta) = I_f(\tau^{(n_k)})$ for all k (outside a P–null set). Taking limits as $k \to \infty$, we get (7.56). ■

Remark 7.2 Note that in (7.57) we have made use of the relation

$$
\int_s^t gf(u)dB(u) = g \int_s^t f(u)dB(u)
$$

for g a bounded $\mathcal{F}_s$–measurable random variable. This obviously holds for nonanticipative step functionals f and, therefore, for all $f \in \mathcal{M}[s, t]$.

Our next task is to extend the notion of a stochastic integral to a larger class of summands than $\mathcal{M}[\alpha, \beta]$.

Definition 7.5 The set of all progressively measurable stochastic processes f satisfying

$$
P\left(\int_\alpha^\beta f^2(s)ds < \infty\right) = 1 \tag{7.58}
$$

is denoted by $\mathcal{L}[\alpha, \beta]$. If (7.58) holds $\forall\, \beta > \alpha$, then this space is $\mathcal{L}[\alpha, \infty)$.

We can proceed with the definition of $\int_\alpha^t f(s)dB(s)$ for $f \in \mathcal{L}[\alpha, \infty)$ as follows: Define the $\{\mathcal{F}_t : \alpha \leq t \leq \beta\}$–stopping times

$$
\tau^{(n)}(\omega) = \inf\{t \geq \alpha : \int_\alpha^t f^2(s, \omega)ds \geq n\} \wedge n, \quad (\text{integer } n \geq \alpha).
$$

Then $\tau^{(n)} \uparrow \infty$ a.s. as $n \uparrow \infty$. (If, for some ω, $\int_\alpha^t f^2(s,\omega)ds < n\ \forall\, t$, then $\tau^{(n)}(\omega) = \infty \wedge n = n$; if $\int_\alpha^t f^2(s,\omega)ds \uparrow \infty$ as $t \uparrow \infty$, then $\inf\{t \geq \alpha : \int_\alpha^t f^2(s,\omega)ds \geq n\} \uparrow \infty$ as $n \uparrow \infty$). Let

$$f_n(s) := f(s)\mathbf{1}_{[\tau^{(n)} \geq s]} \qquad (n \geq \alpha, n \text{ integer}).$$

Then

$$\int_\alpha^T f_n^2(s)ds \leq \int_\alpha^{\tau^{(n)}(\omega)} f^2(s,\omega)ds \leq n.$$

Hence $f_n \in \mathcal{M}[\alpha,\infty)$ and, applying Proposition 7.6 with $\tau = \tau^{(m)} \wedge t$ for $m < n$, and noting that $f_n(t) = f_m(t) = f(t)$ for $t \leq \tau^{(m)}$, one has

$$\begin{aligned} I_{f_n}(t \wedge \tau^{(m)}) &= \int_\alpha^t f_n(s)\mathbf{1}_{[\tau^{(m)} \geq s]}dB_s = \int_\alpha^t f(s)\mathbf{1}_{[\tau^{(m)} \geq s]}dB_s \\ &\equiv I_{f_m}(t) \ \text{ a.s.,} \end{aligned} \tag{7.59}$$

so that, using continuity in t for all terms in (7.59),

$$I_{f_n}(t) = I_{f_m}(t) \quad \forall\, t \leq \tau^{(m)}, \quad \text{a.s., if } n > m. \tag{7.60}$$

Definition 7.6 Let $f \in \mathcal{L}[\alpha,\infty)$. The *stochastic integral*

$$I_f(t) = \int_\alpha^t f(s)dB_s, \alpha \leq s < \infty,$$

is defined by

$$I_f(t) := I_{f_m}(t) \quad \forall\, t \leq \tau^{(m)} \quad (m = [\alpha]+1, [\alpha]+2, \ldots). \tag{7.61}$$

If $f \in \mathcal{L}[\alpha,\beta]$, then the same construction holds for $\alpha \leq t \leq \beta$, with $\tau^{(n)}(\omega) = \inf\{t \in [\alpha,\beta] : \int_\alpha^t f^2(s,\omega)ds \geq n\}$, and $\tau^{(n)}(\omega) = \beta$ if $\int_\alpha^t f^2(s,\omega)ds < n$ for all $t \leq \beta$.

Remark 7.3 Note that $t \to I_f(t)$ is continuous outside a P–null set if $f \in \mathcal{L}[\alpha,\infty)$. With the above definition of the stochastic integral, Proposition 7.6 may be extended to $f \in \mathcal{L}[\alpha,\infty)$ and bounded stopping times $\tau \leq \beta$. To see this note that, with f_n as defined above,

$$\begin{aligned} I_f(\tau) &= \lim_{n\to\infty} I_{f_n}(\tau) = \lim_{n\to\infty} \int_\alpha^\beta f_n(s)\mathbf{1}_{[\tau \geq s]}dB_s \\ &\equiv \lim_{n\to\infty} \int_\alpha^\beta f(s)\mathbf{1}_{[\tau \geq s]}\mathbf{1}_{[\tau^{(n)} \geq s]}dB_s = \lim_{n\to\infty} I_g(\beta \wedge \tau^{(n)}), \end{aligned}$$

where $g(s) := f(s)\mathbf{1}_{[\tau \geq s]}$. We used Proposition 7.6 in the second and the last equalities above. Now let $n \uparrow \infty$ and use the continuity of the stochastic integral to get $I_f(\tau) = I_g(\beta)$.

Proposition 7.6 and its extension given in Remark 7.3 allow us to construct diffusions with *locally Lipschitz* coefficients (parameters); recall (7.54). Assume that $\boldsymbol{\mu}(\cdot), \boldsymbol{\sigma}(\cdot)$ are locally Lipschitzian. One may still construct a diffusion on $\mathbb{R}^k$ satisfying (7.34) up to a random time, ζ, referred to as *explosion time*, as suggested by the definition to follow below. For this, let $\{\mathbf{X}^{\mathbf{x}}_{t,n} : t \geq 0\}$ be the solution of (7.34) with *globally Lipschitzian* coefficients $\boldsymbol{\mu}_n(\cdot), \boldsymbol{\sigma}_n(\cdot)$ satisfying

$$\boldsymbol{\mu}_n(\mathbf{y}) = \boldsymbol{\mu}(\mathbf{y}) \quad \text{and} \quad \boldsymbol{\sigma}_n(\mathbf{y}) = \boldsymbol{\sigma}(\mathbf{y}) \qquad \text{for } |\mathbf{y}| \leq n. \tag{7.62}$$

We may, for example, let $\boldsymbol{\mu}_n(\mathbf{y}) = \boldsymbol{\mu}(n\mathbf{y}/|\mathbf{y}|)$ and $\boldsymbol{\sigma}_n(\mathbf{y}) = \boldsymbol{\sigma}(n\mathbf{y}/|\mathbf{y}|)$ for $|\mathbf{y}| > n$, so that (7.33) holds for *all* $\mathbf{x}, \mathbf{y}$ with $M = M_n$. Let us show that, if $|\mathbf{x}| \leq n$,

$$\mathbf{X}^{\mathbf{x}}_{t,m}(\omega) = \mathbf{X}^{\mathbf{x}}_{t,n}(\omega) \qquad \text{for } 0 \leq t < \zeta_n(\omega) \quad (m \geq n) \tag{7.63}$$

outside a set of zero probability, where

$$\zeta_n(\omega) := \inf\{t \geq 0 : |\mathbf{X}^{\mathbf{x}}_{t,n}(\omega)| = n\}. \tag{7.64}$$

To prove (7.63), first note that if $m \geq n$ then

$$\boldsymbol{\mu}_m(\mathbf{y}) = \boldsymbol{\mu}_n(\mathbf{y}) \text{ and } \boldsymbol{\sigma}_m(\mathbf{y}) = \boldsymbol{\sigma}_n(\mathbf{y}), \text{ for } |\mathbf{y}| \leq n,$$

so that for $0 \leq t < \zeta_n(\omega)$,

$$\boldsymbol{\mu}_m(\mathbf{X}^{\mathbf{x}}_{t,n}) = \boldsymbol{\mu}_n(\mathbf{X}^{\mathbf{x}}_{t,n}) \quad \text{and} \quad \boldsymbol{\sigma}_m(\mathbf{X}^{\mathbf{x}}_{t,n}) = \boldsymbol{\sigma}_n(\mathbf{X}^{\mathbf{x}}_{t,n}). \tag{7.65}$$

It follows from (7.65), Proposition 7.6 with $t \wedge \zeta_n$ for τ and continuity of the stochastic integral, that for $0 \leq t \leq \zeta_n$,

$$\int_0^t \boldsymbol{\mu}_m(\mathbf{X}^{\mathbf{x}}_{s,n})ds + \int_0^t \boldsymbol{\sigma}_m(\mathbf{X}^{\mathbf{x}}_{s,n})d\mathbf{B}_s = \int_0^t \boldsymbol{\mu}_n(\mathbf{X}^{\mathbf{x}}_{s,n})ds + \int_0^t \boldsymbol{\sigma}_n(\mathbf{X}^{\mathbf{x}}_{s,n})d\mathbf{B}_s$$

almost surely. Therefore, a.s. for $0 \leq t \leq \zeta_n$,

$$\mathbf{X}^{\mathbf{x}}_{t,m} - \mathbf{X}^{\mathbf{x}}_{t,n} = \int_0^t [\boldsymbol{\mu}_m(\mathbf{X}^{\mathbf{x}}_{s,m}) - \boldsymbol{\mu}_m(\mathbf{X}^{\mathbf{x}}_{s,n})]ds + \int_0^t [\boldsymbol{\sigma}_m(\mathbf{X}^{\mathbf{x}}_{s,m}) - \boldsymbol{\sigma}_m(\mathbf{X}^{\mathbf{x}}_{s,n})]d\mathbf{B}_u. \tag{7.66}$$

In particular, from sample path continuity, it follows that $\zeta_n, n \geq 1$, is an a.s. monotone sequence so that $\zeta_n \uparrow$ a.s. to (the *explosion time*) ζ, say, as $n \uparrow \infty$, and

$$\mathbf{X}_t^{\mathbf{x}}(\omega) = \lim_{n\to\infty} \mathbf{X}_{t,n}^{\mathbf{x}}(\omega), \qquad t < \zeta(\omega), \tag{7.67}$$

exists a.s., and has a.s. continuous sample paths on $[0, \zeta(\omega))$. If $P(\zeta < \infty) > 0$, then we say that *explosion occurs*. In this case, one may continue the process $\{\mathbf{X}_t^{\mathbf{x}}\}$ for $t \geq \zeta(\omega)$ by setting

$$\mathbf{X}_t^{\mathbf{x}}(\omega) = \text{“}\infty\text{”}, \qquad t \geq \zeta(\omega), \tag{7.68}$$

where “∞” is a new state, usually taken to be the *point at infinity* in the *one-point compactification* of $\mathbb{R}^k$. In addition to all open subsets of $\mathbb{R}^k$, complements of all compact subsets of $\mathbb{R}^k$ with the new point “∞” adjoined to them define the topology of $\mathbb{R}^k \cup \{\text{“}\infty\text{”}\}$. The two points $+\infty$ and $-\infty$ may be used in one dimension.

Using the Markov property of $\{\mathbf{X}_{t,n}^{\mathbf{x}}\}$ for each n, it is not difficult to show that $\{\mathbf{X}_t^{\mathbf{x}} : t \geq 0\}$ defined by (7.67) and (7.68) is a Markov process on $\mathbb{R}^k \cup \{\text{“}\infty\text{”}\}$, called the *minimal diffusion generated by* (Exercise 9)

$$\mathbf{A} = \frac{1}{2} \sum d_{ij}(\mathbf{x}) \frac{\partial^2}{\partial x^{(i)} \partial x^{(j)}} + \sum \mu^{(i)}(\mathbf{x}) \frac{\partial}{\partial x^{(i)}}, \tag{7.69}$$

with *drift vector* $\boldsymbol{\mu}(\cdot)$ and *diffusion matrix* $d(\cdot) = \sigma(\cdot)\sigma'(\cdot)$.

7.4 Strong Markov Property

In this sub-section we derive the important result that the diffusion $\{\mathbf{X}^{\mathbf{x}}, \mathbf{x} \in \mathbb{R}^k\}$ has the strong Markov property. We will denote the distribution of $\mathbf{X}^{\mathbf{x}}$ by Q^x.

Theorem 7.7 *Under the hypothesis of Theorem 7.5, $\mathbf{X}^{\mathbf{x}}$ $(\mathbf{x} \in \mathbb{R}^k)$ has the strong Markov property with respect to $\mathcal{F}_\tau$, for every $\{\mathcal{F}_t : t \geq 0\}$–stopping time τ.*

According to Theorem 1.6, we only need to show the *Feller property*, i.e., the map $\mathbf{x} \to p(t; \mathbf{x}, d\mathbf{y})$ is weakly continuous. The following proposition, which is of independent interest, implies this.

Proposition 7.8 *Under the hypothesis of Theorem 7.5, one has*

$$\mathbb{E}(\max_{0\leq s\leq t} |\mathbf{X}_s^{\mathbf{x}} - \mathbf{X}_s^{\mathbf{z}}|^2) \tag{7.70}$$

$$\leq 3|\mathbf{x} - \mathbf{z}|^2[1 + (3t + 12)M^2 t \exp\{\frac{3M^2t^2}{2} + 12M^2 t\}], \quad 0 \leq t \leq T,\ T > 0,$$

with the matrix norm in (7.49) taken as $\|\mathbf{C}\| = (Trace\ C'C)^{\frac{1}{2}}$.

Proof Since

$$\mathbf{X}_t^{\mathbf{x}} - \mathbf{X}_t^{\mathbf{z}} = \mathbf{x} - \mathbf{z} + \int_0^t \{\boldsymbol{\mu}(\mathbf{X}_s^{\mathbf{x}}) - \boldsymbol{\mu}(\mathbf{X}_s^{\mathbf{z}})\}ds + \int_0^t \{\boldsymbol{\sigma}(\mathbf{X}_s^{\mathbf{x}}) - \boldsymbol{\sigma}(\mathbf{X}_s^{\mathbf{z}})\}d\mathbf{B}_s, \ t \geq 0,$$

one has

$$\begin{aligned}\max_{0\leq s\leq t} |\mathbf{X}_s^{\mathbf{x}} - \mathbf{X}_s^{\mathbf{z}}| \leq |\mathbf{x} - \mathbf{z}| + M \int_0^t |\mathbf{X}_s^{\mathbf{x}} - \mathbf{X}_s^{\mathbf{z}}|ds \\ + \max_{0\leq s\leq t} \left| \int_0^s \{\boldsymbol{\sigma}(\mathbf{X}_u^{\mathbf{x}}) - \boldsymbol{\sigma}(\mathbf{X}_u^{\mathbf{z}})\}d\mathbf{B}_u \right|, \quad t \geq 0.\end{aligned}$$

Using the Cauchy–Schwarz inequality on the middle term and Doob's maximal inequality (for second moment) for the third, one gets

$$\begin{aligned}D_t &:= \mathbb{E}(\max_{0\leq s\leq t} |\mathbf{X}_s^{\mathbf{x}} - \mathbf{X}_s^{\mathbf{z}}|^2) \\ &\leq 3|\mathbf{x} - \mathbf{z}|^2 + 3M^2 t \int_0^t \mathbb{E}|\mathbf{X}_s^{\mathbf{x}} - \mathbf{X}_s^{\mathbf{z}}|^2 ds + 12M^2 \int_0^t \mathbb{E}|\mathbf{X}_s^{\mathbf{x}} - \mathbf{X}_s^{\mathbf{z}}|^2 ds \\ &\leq 3|\mathbf{x} - \mathbf{z}|^2 + (3M^2 t + 12M^2) \int_0^t D_s ds, \quad t \geq 0.\end{aligned}$$

We have here used the relations

$$\begin{aligned}&\mathbb{E}(\max_{0\leq s\leq t} \left| \int_0^s (\boldsymbol{\sigma}(\mathbf{X}_u^{\mathbf{x}}) - \boldsymbol{\sigma}(\mathbf{X}_u^{\mathbf{z}}))d\mathbf{B}_u \right|^2) \\ &= \mathbb{E}(\max_{0\leq s\leq t} \sum_{i=1}^k (\int_0^s (\boldsymbol{\sigma}_i \cdot (\mathbf{X}_u^{\mathbf{x}}) - \boldsymbol{\sigma}_i(\mathbf{X}_u^{\mathbf{z}})) \cdot d\mathbf{B}_u]^2) \\ &\leq \sum_{i=1}^k 4 \int_0^t \mathbb{E}|\boldsymbol{\sigma}_i(\mathbf{X}_u^{\mathbf{x}}) - \boldsymbol{\sigma}_i(X_u^{\mathbf{z}})|^2 du \\ &= 4 \int_0^t \mathbb{E}\|\boldsymbol{\sigma}(\mathbf{X}_u^{\mathbf{x}}) - \boldsymbol{\sigma}(\mathbf{X}_u^{\mathbf{z}})\|^2 du \\ &\leq 4M^2 \int_0^t \mathbb{E}|\mathbf{X}_u^{\mathbf{x}} - \mathbf{X}_u^{\mathbf{z}}|^2 du.\end{aligned}$$

The inequality (7.70) now follows from Gronwall's inequality below. ■

Proof of Theorem 7.7 It follows from (7.70) that $\mathbf{X}^{\mathbf{x}}$ converges to $\mathbf{X}^{\mathbf{z}}$ in probability as $\mathbf{x} \to \mathbf{z}$, with respect to the topology of uniform convergence on compact intervals in the space $C([0, \infty) : \mathbb{R}^k)$. Hence $\mathbf{X}^{\mathbf{x}}$ converges in distribution to $\mathbf{X}^{\mathbf{z}}$ as $\mathbf{x} \to \mathbf{z}$. That is, $\mathbf{x} \to Q^{\mathbf{x}}$ is weakly continuous. ■

Lemma 1 (Gronwall's Inequality) *Suppose an integrable* $\alpha : [0, T] \to \mathbb{R}$ *satisfies the inequality*

$$\alpha(t) \leq C_0 + H(t) \int_0^t \alpha(s)ds, \qquad 0 \leq t \leq T, \tag{7.71}$$

where H *is a nonnegative integrable function on* $[0, T]$*. Then*

$$\alpha(t) \leq C_0[1 + H(t) \int_0^t \exp\{\int_s^t H(u)du\}ds], \qquad 0 \leq t \leq T. \tag{7.72}$$

Proof Write $g(t) = \int_0^t \alpha(s)ds$. Then the inequality (7.71) may be expressed as $g'(t) - H(t)g(t) \leq C_0$ or $(d/dt)[\exp\{-\int_0^t H(s)ds\}g(t)] \leq C_0 \exp\{-\int_0^t H(s)ds\}$. Integrating this one gets

$$g(t) \leq C_0 e^{\int_0^t H(s)ds} \int_0^t e^{-\int_0^s H(u)du} ds = C_0 \int_0^t \exp\{\int_s^t H(u)du\}ds.$$

Using this estimate in the right side of (7.71), one gets (7.72). ■

If one uses the matrix norm

$$\|\mathbf{C}\| := \max_{|\mathbf{x}|=1} |\mathbf{C}\mathbf{x}| = (\text{largest eigenvalue of } \mathbf{C}'\mathbf{C})^{\frac{1}{2}}$$

(under the assumption (7.49)), then in k dimensions the right side of (7.70) changes to $3|\mathbf{x} - \mathbf{z}|^2[1 + (3t + 12k)M^2 t \exp\{\frac{3M^2t^2}{2} + 12M^2kt\}]$ (Exercise 12(b)).

As noted in the previous subsection on processes with locally Lipschitz coefficients, one may establish the Markov property for the minimal process. In fact, one may prove the strong Markov property for the minimal process as well (Exercise 9). Criteria for explosion/nonexplosion based on on the coefficients of the diffusion are provided in Chapter 12.

7.5 An Extension to SDEs with Nonhomogeneous Coefficients

The construction of (nonhomogeneous) diffusions with *time-dependent* drift and diffusion coefficients, say $\boldsymbol{\mu}(\mathbf{x}, t)$ and $\boldsymbol{\sigma}(\mathbf{x}, t)$, will be solved by viewing these as "homogeneous coefficients" for a space-time process $(\mathbf{X}_t, t)$ evolving in $\mathbb{R}^k \times [0, \infty)$. So suppose that we are given functions $\mu^{(i)}(\mathbf{x}, t)$, $\sigma_{ij}(\mathbf{x}, t)$ on $\mathbb{R}^k \times [0, \infty)$ into $\mathbb{R}$. Write $\boldsymbol{\mu}(\mathbf{x}, t)$ for the vector whose ith component is $\mu^{(i)}(\mathbf{x}, t)$, $\boldsymbol{\sigma}(\mathbf{x}, t)$ for the $k \times k$ matrix whose (i, j) element is $\sigma_{ij}(\mathbf{x}, t)$. Given $s \geq 0$, we would like to solve the stochastic differential equation

$$d\mathbf{X}_t = \boldsymbol{\mu}(\mathbf{X}_t, t)dt + \boldsymbol{\sigma}(\mathbf{X}_t, t)d\mathbf{B}_t \qquad (t \geq s), \qquad \mathbf{X}_s = \mathbf{Z}, \tag{7.73}$$

where $\mathbf{Z}$ is $\mathcal{F}_s$-measurable, $\mathbb{E}|\mathbf{Z}|^2 < \infty$. It is not difficult to extend the proofs of Theorems 7.1, 7.2 (also see Theorems 7.4, 7.5) to prove the following result (Exercise 10).

Theorem 7.9 *Assume*

$$\begin{aligned} |\boldsymbol{\mu}(\mathbf{x}, t) - \boldsymbol{\mu}(\mathbf{y}, s)| &\leq M\{|\mathbf{x} - \mathbf{y}| + |t - s|\}, \\ \|\boldsymbol{\sigma}(\mathbf{x}, t) - \boldsymbol{\sigma}(\mathbf{y}, s)\| &\leq M\{|\mathbf{x} - \mathbf{y}| + |t - s|\}, \end{aligned} \tag{7.74}$$

for some constant M.

a. *Then for every $\mathcal{F}_s$-measurable k-dimensional square integrable $\mathbf{Z}$, there exists a unique nonanticipative continuous solution $\mathbf{X} = \{\mathbf{X}_t : t \geq s\}$ in $\mathcal{M}[s, \infty)$ of*

$$\mathbf{X}_t = \mathbf{Z} + \int_s^t \boldsymbol{\mu}(\mathbf{X}_u, u)du + \int_s^t \boldsymbol{\sigma}(\mathbf{X}_u, u)d\mathbf{B}_u \qquad (t \geq s). \tag{7.75}$$

b. *Let $\{\mathbf{X}_t^{s,\mathbf{x}} : t \geq s\}$ denote the solution of (7.75) with $\mathbf{Z} = \mathbf{x} \in \mathbb{R}^k$. Then it is a (generally nonhomogeneous) Markov process with (initial) state $\mathbf{x}$ at time s and transition probability*

$$p(s', t; \mathbf{z}, B) = P(\mathbf{X}_t^{s',\mathbf{z}} \in B) \qquad (s \leq s' \leq t).$$

■

Remark 7.4 To solve (7.73), one may introduce an additional coordinate t and solve the augmented equation for $Y_t = (\mathbf{X}_t, t)'$, with a drift $(\boldsymbol{\mu}(\mathbf{x}, t), 1)'$, diffusion matrix $(\boldsymbol{\sigma}(\mathbf{x}, t), \mathbf{0})$ augmented by a column of zeros, and a $(k + 1)$-dimensional Brownian motion $(\mathbf{B}_t, B_t^{(k+1)})$, $t \geq 0$.

7.6 An Extension to k−Dimensional SDE Governed by r−Dimensional Brownian Motion

Finally, we consider k–dimensional diffusions $\mathbf{X}_t = (X_t^{(1)}, \ldots, X_t^{(k)})'$ governed by r-dimensional Brownian motions $\mathbf{B}_t = (B_t^{(1)}, \ldots, B_t^{(r)})$ illustrated by the following examples.

Example 3 (k = 1, r =2) Suppose that $B_j = \{B_j(t) : t \geq 0\}$, $j = 1, 2$, are two independent one-dimensional Brownian motions and one wishes to consider the one-dimensional diffusion defined by

$$dX(t) = \sigma_1 dB_1(t) + \sigma_2 dB_2(t), \tag{7.76}$$

for positive constants σ_1, σ_2. Then, up to specification of the initial state, $X = \{X(t) : t \geq 0\}$ is well-defined and equivalently given by the one-dimensional diffusion equation

$$dX(t) = \sqrt{\sigma_1^2 + \sigma_2^2} dB(t). \tag{7.77}$$

The following example illustrates another possible dimensional effect.

Example 4 (k = 2, r = 1) Suppose that $\{B_t : t \geq 0\}$ is a one-dimensional standard Brownian motion and consider the two-dimensional diffusion $\{(X_1(t), X_2(t)) : t \geq 0\}$ defined by

$$\begin{aligned} dX_1(t) &= \sigma_1 dB_t \\ dX_2(t) &= \sigma_2 dB_t. \end{aligned} \tag{7.78}$$

In this case, the "two-dimensional" process is also well defined and is determined by a single one-dimensional Brownian motion.

More generally, let $\boldsymbol{\mu}(\cdot)$ be a Lipschitz function on $\mathbb{R}^k$ into $\mathbb{R}^k$ and $\boldsymbol{\sigma}(\cdot)$ a Lipschitz function on $\mathbb{R}^k$ into the space of $k \times r$ matrices. Here r need not equal k. Consider now the Equations (7.50), (7.51) with these changes. In particular, $\sigma_{i\cdot}(\mathbf{x}) := (\sigma_{i1}(\mathbf{x}), \cdots, \sigma_{ir}(\mathbf{x}))$ is the i-th row of $\boldsymbol{\sigma}(\mathbf{x})$, $1 \leq i \leq k$:

$$X_t^{(i)} = X_\alpha^{(i)} + \int_\alpha^t \mu^{(i)}(\mathbf{X}_s)ds + \int_\alpha^t \sum_{j=1}^r \sigma_{ij}(\mathbf{X}_s)dB_s^{(j)}, \quad (1 \leq i \leq k).$$

The proofs of Theorems 7.4, 7.5 go over without any essential change (Exercise 11).

Exercises

1. If h is a Lipschitzian function on $\mathbb{R}$ and X is a square integrable random variable, then prove that $Eh^2(X) < \infty$. [*Hint:* Look at $h(X) - h(0)$.]
2. Assume the hypothesis of Theorem 7.1 with $\alpha = 0$.

   (i) Prove that

   $$\mathbb{E}(X_t - X_0)^2 \leq 4M^2(t+1)\int_0^t \mathbb{E}(X_s - X_0)^2 ds + 4t^2\mathbb{E}\mu^2(X_0) + 4t\mathbb{E}\sigma^2(X_0).$$

   (ii) Write $D_t := \mathbb{E}(\max\{(X_s - X_0)^2 : 0 \leq s \leq t\})$, $0 \leq t \leq T$. Show from (i) that

(a) $D_t \le d_1 \int_0^t D_s ds + d_2$, where $d_1 = 4M^2(t+4)$, $d_2 = 4t^2 \mathbb{E}\mu^2(X_0) + 16t\mathbb{E}\sigma^2(X_0))$;
(b) $D_T \le d_2 e^{d_1 T}$.

3. Write out a proof of Theorem 7.4 by extending step by step the proof of Theorem 7.1.
4. Write out a proof of Theorem 7.5 along the lines of that of Theorem 7.2.
5. Calculate the bound on $D_T^{(1)}$ in (7.14).
6. Show that measurability of $\theta(s,t;z,B_s^t)$ follows from (7.5)–(7.7).
7. Consider the Gaussian diffusion $X^x = \{X_t^x : t \ge 0\}$ given by (7.36).
   (i) Show that $\mathbb{E}X_t^x = e^{-t\gamma}x$.
   (ii) Compute $\mathrm{Cov}(X_t^x, X_{t+h}^x)$.
   (iii) Compute the transition probability distribution $p(t;x,dy)$ for the Ornstein-Uhlenbeck process;[*Hint:* The finite dimensional distributions have a Gaussian density.]
   (iv) For the case $\gamma > 0$, $\sigma \neq 0$, compute the asymptotic distribution of X_t as $t \to \infty$. [*Hint:* The Riemann integral $\int_0^t e^{-\gamma(t-s)} B_s ds$ is a limit of partial sums, and each partial sum has a Gaussian distribution.]
8. (i) Prove (7.30), and (7.32), under the assumption (7.3). [*Hint:* Show that the distribution of $X_{t+h} - X_t$, given $X_t = x$, differs from the distribution of $\mu(x)h + \sigma(x)B_h$ by $o(h)$ as $h \downarrow 0$, where $B_h = B_{t+h} - B_h$ is Gaussian with mean zero and variance h. Use the expected squared error between the two quantities.]
   (ii) Write out a corresponding proof for the multidimensional case.
9. Show that the minimal process $\mathbf{X}^{\mathbf{x}}$ defined by (7.67), (7.68) on $\mathbb{R}^k \cup \{\infty\}$ is Markov and strong Markov.
10. (a) Write out a detailed proof of Theorem 7.9 along the lines of those of Theorems 7.1, 7.2.
    (b) Show that a different proof of Theorem 7.9 follows directly from Theorems 7.4, 7.5, using Remark 7.4, assuming $|\boldsymbol{\mu}(\mathbf{x},t) - \boldsymbol{\mu}(\mathbf{y},s)| \le M(|\mathbf{x}-\mathbf{y}| + |t-s|)$, $|\boldsymbol{\sigma}(\mathbf{x},t) - \boldsymbol{\sigma}(\mathbf{y},s)| \le M(|\mathbf{x}-\mathbf{y}| + |t-s|)$.
11. Follow through in verifying that the proofs of existence and uniqueness of solutions and their Markov property remain essentially unchanged for rectangular systems.
12. (a) (i) Suppose (7.49) holds with respect to the matrix norm $\|\mathbf{C}\| := (\text{Trace } \mathbf{C}'\mathbf{C})^{\frac{1}{2}}$. Then prove that the estimates of $D_T^{(n+1)}$ in k-dimension ($k > 1$) satisfy (7.13) (with c_1 as in (7.14)), and $D_T^{(1)}$ satisfies (7.14) with $\mu^2(\cdot)$ and $\sigma^2(\cdot)$ replaced by $|\boldsymbol{\mu}(\cdot)|^2$ and $\|\boldsymbol{\sigma}(\cdot)\|^2$, respectively.
    (ii) How do these estimates change if the matrix norm used is $\|\mathbf{C}\| = \max_{|\mathbf{x}|=1} |\mathbf{C}\mathbf{x}| = (\text{largest eigenvalue of } \mathbf{C}'\mathbf{C})^{\frac{1}{2}}$?
    (b) Show that with the matrix norm as in (a)(ii) (used in assumption (7.33)), the right side of (7.70) needs to be modified as indicated following the proof of Gronwall's inequality (Lemma 1).

13. Let A and $\boldsymbol{\sigma}$ be a $k \times k$ real matrices and use Picard iteration to solve the the k-dimensional Langevin equation

$$d\mathbf{X}(t) = A\mathbf{X}(t)dt + \boldsymbol{\sigma} d\mathbf{B}(t), t > 0, \quad \mathbf{X}(0) = \mathbf{x}.$$

Under what conditions is the solution a nondegenerate k-dimensional Gaussian process?

Chapter 8
Itô's Lemma

This chapter includes one of the most impactful "chain rule" formulae of the Itô calculus, the celebrated Itô's lemma. Both standard and novel illustrative applications are included. In addition, well-known and useful criteria for recurrence and transience of one-dimensional diffusions are derived as an elegant consequence of Itô's theory.

As observed in Proposition 6.1 and Exercises 1, 2 in Chapter 6, the sample paths of Brownian motion are nowhere differentiable. By Proposition 6.1, it follows that a.s.,

$$\max_{1\le j<k\le 2^n} \left|\sum_{m=j}^{k-1}(B_{(m+1)2^{-n}t} - B_{m2^{-n}t})^2 - (k-j)2^{-n}t\right| \to 0. \tag{8.1}$$

In particular,

$$\max_{0\le m\le 2^n-1} |(B_{(m+1)2^{-n}t} - B_{m2^{-n}t})^2 - 2^{-n}t| \to 0 \quad \text{a.s.} \qquad \text{as } n\to\infty. \tag{8.2}$$

So the quadratic variation of $B = \{B_t : t \ge 0\}$ over an interval is precisely the length of the interval. This provides the justification for a distinctive and extremely important "chain rule" for the stochastic calculus, called *Itô's lemma*. In particular, the proof of Itô's lemma will reveal that (6.4) is the rigorous basis for the often used formalism

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_8

$$(dB_t)^2 = (B_{t+dt} - B_t)^2 = dt$$

in stochastic calculus. To see the use of this formalism in suggesting Itô's chain rule, suppose that $X^x = \{X_t^x : t \geq 0\}$ is a diffusion on $\mathbb{R}$ with Lipschitz coefficients $\mu(\cdot)$ and $\sigma(\cdot)$. Then for a real-valued function φ on $\mathbb{R}$ having two continuous and bounded derivatives, one has

$$d\varphi(X_t^x) = A\varphi(X_t^x)dt + \varphi'(X_t^x)dB_t, \tag{8.3}$$

where $A = \mu(x)\frac{d}{dx} + \frac{1}{2}\sigma^2(x)\frac{d^2}{dx^2}$. That is,

$$\varphi(X_t^x) = \varphi(x) + \int_0^t A\varphi(X_s^x)ds + \int_0^t \varphi'(X_s^x)dB_s. \tag{8.4}$$

The argument is as follows: In view of the quadratic variation computation, it follows from a two-term Taylor expansion

$$\begin{aligned}
&d\varphi(X_t^x) \\
&= \varphi'(X_t^x)dX_t^x + \frac{1}{2}\varphi''(X_t^x)(dX_t^x)^2 + o(dt) \\
&= \varphi'(X_t^x)(\mu(X_t^x)dt + \sigma(X_t^x)dB_t) + \frac{1}{2}\varphi''(X_t^x)(\sigma^2(X_t^x)(dB_t)^2 + o(dt)) + o(dt) \\
&= A\varphi(X_t^x)dt + \varphi'(X_t^x)\sigma(X_t^x)dB_t + o(dt).
\end{aligned} \tag{8.5}$$

For the case of a k-dimensional diffusion $\mathbf{X}^x$ defined by

$$d\mathbf{X}_t^{\mathbf{x}} = \mu(\mathbf{X}_t^{\mathbf{x}})dt + \sigma(\mathbf{X}_t^{\mathbf{x}})d\mathbf{B}_t, \tag{8.6}$$

a similar argument for a twice continuously differentiable real-valued function φ on $\mathbb{R}^k$ with bounded derivatives is possible, leading to

$$d\varphi(\mathbf{X}_t^{\mathbf{x}}) = \mathbf{A}\varphi(\mathbf{X}_t^{\mathbf{x}})dt + \nabla\varphi(\mathbf{X}_t^{\mathbf{x}})d\mathbf{B}_t, \tag{8.7}$$

where the form of the infinitesimal generator $\mathbf{A}$ of the diffusion $\mathbf{X}^x$ is given by

$$\mathbf{A} = \sum_{1\leq i\leq k} \mu^{(i)}(\mathbf{x})\frac{\partial}{\partial x^i} + \frac{1}{2}\sum_{1\leq i,j\leq k} d_{ij}(\mathbf{x})\frac{\partial^2}{\partial x^i \partial x^j}, \tag{8.8}$$

for $D = ((d_{ij})) = \sigma\sigma'$. In particular, again using a Taylor expansion to include quadratic terms, one has

$$\begin{aligned}
d\varphi(\mathbf{X}_t^x) &= \varphi(\mathbf{X}_{t+dt}^{\mathbf{x}}) - \varphi(\mathbf{X}_t^{\mathbf{x}}) \\
&= \mathrm{grad}\varphi(\mathbf{X}_t^{\mathbf{x}}) \cdot d\mathbf{X}_t^{\mathbf{x}} + \frac{1}{2} \sum_{1 \le i, j \le k} \frac{\partial \varphi^2}{\partial x^i \partial x^j} (\mathbf{X}_t^{\mathbf{x}})(dX_t^{\mathbf{x}^i})(dX_t^{\mathbf{x}^j}) + o(dt) \\
&= \sum_{1 \le i \le k} \mu^{(i)}(\mathbf{X}_t^{\mathbf{x}}) \frac{\partial \varphi}{\partial x^i} (\mathbf{X}_t^{\mathbf{x}}) dt + \mathrm{grad}\varphi(\mathbf{X}_t^{\mathbf{x}}) \cdot \sigma(\mathbf{X}_t^{\mathbf{x}})) d\mathbf{B}_t \\
&\quad + \frac{1}{2} \sum_{1 \le i \le j \le k} \frac{\partial^2 \varphi}{\partial x^i \partial x^j} (\mathbf{X}_t^{\mathbf{x}}) \sum_{1 \le r \le k} \sigma_{ir}(\mathbf{X}_t^{\mathbf{x}}) dB_t^{(r)} \sum_{1 \le r' \le k} \sigma_{jj'}(\mathbf{X}_t^{\mathbf{x}}) dB_t^{(r')} \\
&\quad + o(dt).
\end{aligned} \tag{8.9}$$

From here, one invokes another key element of the stochastic calculus formalism to complete the derivation, namely,

$$\big(d\mathbf{B}_t^{(r)}\big)^2 = dt, \quad \big(d\mathbf{B}_t^{(r)}\big)\big(d\mathbf{B}_t^{(r')}\big) = o(dt), r \neq r'.$$

The "smaller order correlation"of increments can be seen as follows.

Proposition 8.1 *Let* $\mathbf{B} = \{\mathbf{B}_t : t \ge 0\}$ *be a k-dimensional standard Brownian motion starting at zero. Fix* $t > 0$. *Then one has for* $i \neq j$, *as* $n \to \infty$ *a.s.,*

$$\max_{1 \le k \le 2^n} \Big| \sum_{m=0}^{k-1} \big(B^{(i)}_{(m+1)2^{-n}t} - B^{(i)}_{m2^{-n}t}\big)\big(B^{(j)}_{(m+1)2^{-n}t} - B^{(j)}_{m2^{-n}t}\big)\Big| \to 0. \tag{8.10}$$

Proof Note that the $2^n + 1$ terms $(B^{(i)}_{(m+1)2^{-n}t} - B^{(i)}_{m2^{-n}t})(B^{(j)}_{(m+1)2^{-n}t} - B^{(j)}_{m2^{-n}t})$, $m = 0, 1, \ldots, 2^n = N$ are independent with common mean zero. Therefore, summing them up to $k \le 2^n$ defines a martingale sequence $\tilde{Y}_k \equiv \tilde{Y}_{k,n}$, for $k = 1, \ldots, N = 2^n$. Applying Doob's maximal inequality, one obtains

$$\begin{aligned}
&P\Big(\max_{1 \le k \le 2^n} \Big| \sum_{m=0}^{k-1} \big(B^{(i)}_{(m+1)2^{-n}t} - B^{(i)}_{m2^{-n}t}\big)\big(B^{(j)}_{(m+1)2^{-n}t} - B^{(j)}_{m2^{-n}t}\big)\Big| > 2^{-\frac{n}{4}}\Big) \\
&\le \mathbb{E}(\tilde{Y}_N)^2 / 2^{-\frac{n}{2}} \\
&= N 2^{\frac{n}{2}} \mathbb{E}\big(B^{(i)}_{(m+1)2^{-n}t} - B^{(i)}_{m2^{-n}t}\big)^2 \big(B^{(j)}_{(m+1)2^{-n}t} - B^{(j)}_{m2^{-n}t}\big)^2 \\
&= N(2^{-n}t)^2 2^{\frac{n}{2}} = 2^{-\frac{n}{2}} t^2.
\end{aligned} \tag{8.11}$$

Thus, one may apply the Borel–Cantelli lemma to obtain the assertion. ■

The following are important corollaries to Itô's lemma as formulated in (8.7).

Corollary 8.2 *Suppose that φ is a twice continuously differentiable real-valued function with bounded derivatives on $\mathbb{R}^k$. Let $\mathbf{X}^{\mathbf{x}}$ be the k-dimensional diffusion defined by*

$$d\mathbf{X}_t^{\mathbf{x}} = \mu(\mathbf{X}_t^{\mathbf{x}})dt + \sigma(\mathbf{X}_t^{\mathbf{x}})d\mathbf{B}_t, \tag{8.12}$$

with Lipschitz coefficients $\mu(\mathbf{x}), \sigma(\mathbf{x})$. Then

$$M_t := \varphi(\mathbf{X}_t^{\mathbf{x}}) - \int_0^t \mathbf{A}\varphi(\mathbf{X}_s^{\mathbf{x}})ds, t \geq 0,$$

is a martingale, where

$$\mathbf{A} = \sum_{1\leq i\leq k} \mu^{(i)}(\mathbf{x})\frac{\partial}{\partial x^i} + \frac{1}{2}\sum_{1\leq i,j\leq k} d_{ij}(\mathbf{x})\frac{\partial^2}{\partial x^i \partial x^j}, \tag{8.13}$$

for $D = ((d_{ij})) = \sigma\sigma^t$.

As a direct consequence of the above corollary and the optional stopping theorem, one obtains the following result.

Corollary 8.3 *Under the hypothesis of Corollary 8.2, assume that*

$$\mathbf{A}\varphi(\mathbf{x}) = 0 \textit{ for all } \mathbf{x}.$$

Then for any stopping time τ such that $\{\mathbf{X}_{\tau\wedge t} : t \geq 0\}$ is uniformly integrable, one has

$$\mathbb{E}\varphi(\mathbf{X}_\tau^{\mathbf{x}}) = \varphi(\mathbf{x}),\ x \in \mathbb{R}^k.$$

Definition 8.1 A function $\varphi \in \mathcal{D}_A$ such that $\mathbf{A}\varphi(\mathbf{x}) = 0$ for all $\mathbf{x}$, is said to be $\mathbf{A}$-harmonic.

Example 1 For an example application, consider the evaluation of $\int_0^t B_s dB_s$. Consider $Y(t) = \frac{1}{2}B^2(t)$. Then, by Itô's lemma, noting $dB(t) = 0dt + 1dB_t$,

$$dY(t) = \frac{1}{2}dt + B_t dB_t.$$

It follows that

$$\int_0^t B_s dB_s = Y(t) - \frac{1}{2}t = \frac{1}{2}B_t^2 - \frac{1}{2}t.$$

Similarly, one may evaluate $\int_0^t B_s^2 dB_s = \frac{1}{3}B_t^3 - \int_0^t B_s ds$.

Example 2 Using the fact that stochastic integrals are martingales with expectation zero, one may obtain a simple recursion between even moments B_t^{2m} as follows. By Itô's lemma,

$$\begin{aligned}\mathbb{E}B_t^{2m} &= \mathbb{E}(\frac{2m(2m-1)}{2}\int_0^t B_s^{2m-2}ds + \int_0^t 2mB_s^{2m-1}dB_s) \\ &= \frac{2m(2m-1)}{2}\int_0^t \mathbb{E}B_s^{2m-2}ds.\end{aligned}$$

Thus $\mathbb{E}B_t^2 = t, \mathbb{E}B_t^4 = 3t^2, \ldots$.

The same argument as above applied to a function $\varphi(t, y)$ on $[0, T] \times \mathbb{R}$, such that $\varphi_0 := \partial\varphi/\partial t$, φ', φ'' are continuous and bounded, leads to

$$\begin{aligned}d\varphi(t, Y(t)) = \{&\varphi_0(t, Y(t)) + \varphi'(t, Y(t))f(t) + \tfrac{1}{2}\varphi''(t, Y(t))g^2(t)s\}dt \\ &+\varphi'(t, Y(t))g(t)dB_t.\end{aligned} \tag{8.14}$$

To state an extension to multidimensions, let $\varphi(t, \mathbf{y})$ be a real-valued function on $[0, T] \times \mathbb{R}^m$ that is once continuously differentiable in t and twice in $\mathbf{y}$. Write

$$\partial_0\varphi(t, \mathbf{y}) := \frac{\partial\varphi(t, \mathbf{y})}{\partial t}, \qquad \partial_r\varphi(t, \mathbf{y}) := \frac{\partial\varphi(t, \mathbf{y})}{\partial y^{(r)}} \qquad (1 \le r \le m). \tag{8.15}$$

Let $\mathbf{B} = \{\mathbf{B}_t : t \ge 0\}$ be a k-dimensional standard Brownian motion satisfying conditions ((6.9)(i), (ii)) (with B_s replaced by $\mathbf{B}_s$). Suppose $\mathbf{Y}(t)$ is a vector of m processes of the form

$$\begin{aligned}\mathbf{Y}(t) &= (Y^{(1)}(t), \ldots, Y^{(m)}(t)), \\ Y^{(r)}(t) &= Y^{(r)}(0) + \int_0^t f_r(s)ds + \int_0^t \mathbf{g}_r(s) \cdot d\mathbf{B}_s \quad (1 \le r \le m).\end{aligned} \tag{8.16}$$

Here, $f_1, \ldots, f_m$ are real-valued and $\mathbf{g}_1, \ldots, \mathbf{g}_m$ vector-valued (with values in $\mathbb{R}^k$) nonanticipative functionals belonging to $\mathcal{M}[0, T]$. Also $Y^{(r)}(0)$ are $\mathcal{F}_0$-measurable square integrable random variables. One may express (8.16) in the differential form

$$dY^{(r)}(t) = f_r(t)dt + \mathbf{g}_r(t) \cdot d\mathbf{B}_t \qquad (1 \le r \le m), \tag{8.17}$$

or, equivalently, expressing $\mathbf{Y}$ and $\mathbf{f}$ as $m \times 1$ vectors, $\mathbf{g}$ an $m \times k$ matrix with row vectors $\mathbf{g}_r$, $r = 1, \ldots m$, and $\mathbf{B}$ as a $k \times 1$ vector, one may also express the system, suitably interpreted coordinate-wise, as

$$d\mathbf{Y}(t) = \mathbf{f}(t)dt + \mathbf{g}(t)d\mathbf{B}_t, \quad t > 0. \tag{8.18}$$

Itô's lemma says

$$d\varphi(t, \mathbf{Y}(t)) = \{\partial_0\varphi(t, \mathbf{Y}(t)) + \sum_{r=1}^{m} \partial_r\varphi(t, \mathbf{Y}(t)) f_r(t)$$
$$+\frac{1}{2!} \sum_{1\le r,r'\le m} \partial_r\partial_{r'}\varphi(t, \mathbf{Y}(t))(\mathbf{g}_r(t)\cdot \mathbf{g}_{r'}(t))\}dt$$
$$+\sum_{r=1}^{m} \partial_r\varphi(t, \mathbf{Y}(t))\mathbf{g}_r(t)\cdot d\mathbf{B}_t. \tag{8.19}$$

In order to arrive at this, write

$$d\varphi(t, \mathbf{Y}(t)) = \varphi(t+dt, \mathbf{Y}(t+dt)) - \varphi(t, \mathbf{Y}(t)) = \varphi(t+dt, \mathbf{Y}(t+dt))$$
$$-\varphi(t, \mathbf{Y}(t+dt)) + \varphi(t, \mathbf{Y}(t+dt) - \varphi(t, \mathbf{Y}(t))$$
$$= \partial_0\varphi(t, \mathbf{Y}(t))dt + \sum_{r=1}^{m} \partial_r\varphi(t, \mathbf{Y}(t))dY^{(r)}(t)$$
$$+\frac{1}{2!} \sum_{1\le r,r'\le m} \partial_r\partial_{r'}\varphi(t, \mathbf{Y}(t))dY^{(r)}(t)dY^{(r')}(t). \tag{8.20}$$

In the usual Newton–Leibniz calculus, of course, the contribution of the last sum to the differential would be zero. But there is one term in the product $dY^{(r)}(t)dY^{(r')}(t)$, which is of the order dt and, therefore, must be retained in computing the stochastic differential. This term in (see (8.17))

$$(\mathbf{g}_r(t)\cdot d\mathbf{B}_t)(\mathbf{g}_{r'}(t)\cdot d\mathbf{B}_t)$$
$$= (\sum_{i=1}^{k} g_r^{(i)}(t)dB_t^{(i)})(\sum_{j=1}^{k} g_{r'}^{(j)}(t)dB_t^{(j)})$$
$$= \sum_{i=1}^{k} g_r^{(i)}(t)g_{r'}^{(i)}(t)(dB_t^{(i)})^2 + \sum_{i\ne j} g_r^{(i)}(t)g_{r'}^{(j)}(t)dB_t^{(i)}dB_t^{(j)}. \tag{8.21}$$

Now as seen above (see (8.10)), the first sum in (8.21) equals

$$(\sum_{i=1}^{k} g_r^{(i)}(t)g_{r'}^{(i)}(t))dt = \mathbf{g}_r(t)\cdot \mathbf{g}_{r'}(t)dt. \tag{8.22}$$

To show that the contribution of the second term in (8.21) to $\varphi(t, \mathbf{Y}(t)) - \varphi(0, \mathbf{Y}(0))$ over any interval $[0, t]$ is zero, note that, for $i \ne j$,

$$Z'_{k,n} := \sum_{m=0}^{k-1} g_r^{(i)}(m2^{-n}t)g_{r'}^{(j)}(\frac{m}{2^n}t)(B^{(i)}_{\frac{m+1}{2^n}t} - B^{(i)}_{\frac{m}{2^n}t})(B^{(j)}_{\frac{m+1}{2^n}t} - B^{(j)}_{\frac{m}{2^n}t}),$$
$$1 \leq k \leq 2^n, \qquad (8.23)$$

is a martingale. This is true as the conditional expectation of the $m-$th summand, given $\mathcal{F}_{m2^{-n}t}$, is zero. Therefore, as in (8.10),

$$\max_{1\leq k\leq 2^n} |Z'_{k,n}| \to 0 \quad \text{a.s.} \qquad \text{as } n \to \infty. \qquad (8.24)$$

Thus,

$$\sum_{i\neq j} g_r^{(i)}(t)g_{r'}^{(j)}(t)dB_t^{(i)}dB_t^{(j)} = 0 \qquad (i \neq j). \qquad (8.25)$$

Using (8.22), (8.25) in (8.20), Itô's lemma (8.19) is obtained. A more elaborate argument is given below. For ease of reference, here is a statement of Itô's lemma from a perspective of stochastic calculus.

Theorem 8.4 (Itô's Lemma) *Assume $f_1, \ldots, f_m$, $\mathbf{g}_1, \ldots, \mathbf{g}_m$ belong to $\mathcal{L}[0, T]$, and $Y^{(r)}(0)$ $(1 \leq r \leq m)$ are $\mathcal{F}_0$-measurable and square integrable. Let $Y = \{Y(t) : t \geq 0\}$ denote the process defined by (8.16). Let $\varphi(t, \mathbf{y})$ be a real-valued function on $[0, T] \times \mathbb{R}^m$, which is once continuously differentiable in t and twice in $\mathbf{y}$. Then (8.19) holds, i.e., for $0 \leq s < t \leq T$,*

$$\begin{aligned}\varphi(t, \mathbf{Y}(t)) - \varphi(s, \mathbf{Y}(s)) = \int_s^t \{&\partial_0\varphi(u, \mathbf{Y}(u)) + \sum_{r=1}^m \partial_r\varphi(u, \mathbf{Y}(u))f_r(u) \\ &+\frac{1}{2!}\sum_{1\leq r,r'\leq m} \partial_r\partial_{r'}\varphi(u, \mathbf{Y}(u))\mathbf{g}_r(u)\cdot\mathbf{g}_{r'}(u)\}du \\ &+\sum_{r=1}^m\int_s^t \partial_r\varphi(u, \mathbf{Y}(u))\mathbf{g}_r(u)\cdot d\mathbf{B}_u. \end{aligned} \qquad (8.26)$$

Proof Assume first that $f_r, \mathbf{g}_r$ are nonanticipative step functionals bounded by a constant and that $\partial_0\varphi$, $\partial_r\varphi$, $\partial_r\partial_{r'}\varphi$ are bounded on $[0, T] \times \mathbb{R}^m$. The general case follows by approximating $f_r, \mathbf{g}_r$ by step functionals and taking limits in probability.

Fix $s < t$ $(0 \leq s < t \leq T)$. In view of the additivity of the (Riemann and stochastic) integrals, it is enough to consider the case $f_r(s') = f_r(s)$, $\mathbf{g}_r(s') = \mathbf{g}_r(s)$ for $s \leq s' \leq t$. Let $t_0^{(n)} = s < t_1^{(n)} < \cdots < t_{N_n}^{(n)} = t$ be a sequence of partitions such that

$$\delta_n := \max\{t_{i+1}^{(n)} - t_t^{(n)} : 0 \le i \le N_n - 1\} \to 0 \qquad \text{as } n \to \infty.$$

Write

$$\varphi(t, \mathbf{Y}(t)) - \varphi(s, \mathbf{Y}(s)) = \sum_{i=0}^{N_n-1} [\varphi(t_{i+1}^{(n)}, \mathbf{Y}(t_{i+1}^{(n)})) - \varphi(t_i^{(n)}, \mathbf{Y}\big(t_{i+1}^{(n)}))]$$
$$+ \sum_{i=0}^{N_n-1} [\varphi(t_i^{(n)}, \mathbf{Y}(t_{i+1}^{(n)})) - \varphi(t_i^{(n)}, \mathbf{Y}(t_i^{(n)}))]. \quad (8.27)$$

The *first sum* on the right may be expressed as

$$\sum_{i=0}^{N_n-1} (t_{i+1}^{(n)} - t_i^{(n)})\{\partial_0\varphi(t_i^{(n)}, \mathbf{Y}(t_{i+1}^{(n)})) + R_{n,i}^{(1)}\} \quad (8.28)$$

where

$$|R_{n,i}^{(1)}| \le \max\{|\partial_0\varphi(u, \mathbf{Y}(t_{i+1}^{(n)})) - \partial_0\varphi(u', \mathbf{Y}(t_{i+1}^{(n)}))| : t_i^{(n)} \le u \le u' \le t_{i+1}^{(n)}\} \to 0$$

uniformly in i (for each ω), as $n \to \infty$, because $\partial_0\varphi$ is uniformly continuous on $[s, t] \times \{Y(u, \omega) : s \le u \le t\}$. Hence, (8.28) converges, for all ω, to

$$J_0 := \int_s^t \partial_0\varphi(u, \mathbf{Y}(u))du. \quad (8.29)$$

By a Taylor expansion, the *second sum* on the right in (8.27) may be expressed as

$$\sum_{i=0}^{N_n-1} \sum_{r=1}^{m} (Y^{(r)}(t_{i+1}^{(n)}) - Y^{(r)}(t_i^{(n)}))\partial_r\varphi(t_i^{(n)}, \mathbf{Y}(t_i^{(n)}))$$
$$+ \frac{1}{2!} \sum_{1\le r,r'\le m} \sum_{i=0}^{N_n-1} (Y^{(r)}(t_{i+1}^{(n)}) - Y^{(r)}(t_i^{(n)}))(Y^{(r')}(t_{i+1}^{(n)}) - Y^{(r')}(t_i^{(n)}))$$
$$\times\{\partial_r\partial_{r'}\varphi(t_i^{(n)}, \mathbf{Y}(t_i^{(n)})) + R_{r,r',n,i}^{(2)}\}, \quad (8.30)$$

where $R_{r,r',n,i}^{(2)} \to 0$ uniformly in i (for where each ω), because $(u, \mathbf{y}) \to \partial_r\partial_{r'}\varphi(u, \mathbf{y})$ and $t \to \mathbf{Y}(t)$ are continuous. Using (8.16) the in-probability limit in the first sum in (8.30) is expressed as

$$\sum_{r=1}^{m} \sum_{i=0}^{N_n-1} \{(t_{i+1}^{(n)} - t_i^{(n)})f_r(t_i^{(n)}) + \mathbf{g}_r(t_i^{(n)}) \cdot (\mathbf{B}_{t_{i+1}^{(n)}} - \mathbf{B}_{t_i^{(n)}})\}\partial_r\varphi(t_i^{(n)}, \mathbf{Y}(t_i^{(n)}))$$

$$\to J_{11} := \sum_{r=1}^{m} [\int_s^t f_r(u)\partial_r\varphi(u, \mathbf{Y}(u))du + \int_s^t \partial_r\varphi(u, \mathbf{Y}(u))\mathbf{g}_r(u) \cdot d\mathbf{B}_u]. \tag{8.31}$$

For $\partial_r\varphi$ is continuous and bounded, so that

$$\sum_{i=0}^{N_n-1} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (\partial_r\varphi(t_i^{(n)}, \mathbf{Y}(t_i^{(n)})) - \partial_r\varphi(u, \mathbf{Y}(u)))\mathbf{g}_r(u) \cdot d\mathbf{B}_u = \int_s^t h_n(u) \cdot d\mathbf{B}_u \to 0,$$

in probability, since $\int_s^t |h_n(u)|^2 du \to 0$. It remains to find the limiting value of the second sum in (8.30) (excluding the remainder). For this, write

$$\begin{aligned}
&\sum_{i=0}^{N_n-1} \big(Y^{(r)}\big(t_{i+1}^{(n)}\big) - Y^{(r)}\big(t_i^{(n)}\big)\big)\big(Y^{(r')}\big(t_{i+1}^{(n)}\big) - Y^{(r')}\big(t_i^{(n)}\big)\big)\partial_r\partial_{r'}\varphi\big(t_i^{(n)}, \mathbf{Y}\big(t_i^{(n)}\big)\big)\\
&\quad = \sum_{i=0}^{N_n-1} \big\{\big(t_{i+1}^{(n)} - t_i^{(n)}\big) f_r\big(t_i^{(n)}\big) + \mathbf{g}_r\big(t_i^{(n)}\big) \cdot (\mathbf{B}_{t_{i+1}^{(n)}} - \mathbf{B}_{t_i^{(n)}})\big\}\\
&\qquad \times\big\{\big(t_{i+1}^{(n)} - t_i^{(n)}\big) f_{r'}\big(t_i^{(n)}\big) + \mathbf{g}_{r'}\big(t_i^{(n)}\big) \cdot \big(\mathbf{B}_{t_{i+1}^{(n)}} - \mathbf{B}_{t_i^{(n)}}\big)\big\}\partial_r\partial_{r'}\varphi\big(t_i^{(n)}, \mathbf{Y}\big(t_i^{(n)}\big)\big).
\end{aligned} \tag{8.32}$$

By (8.10), (8.23) and (8.24), (8.32) converges a.s. (for a suitable sequence of partitions) to

$$\begin{aligned}
&\lim \sum_{i=0}^{N_n-1} \{\sum_{j=1}^{k} g_r^{(j)}(t_i^{(n)})(B_{t_{i+1}^{(n)}}^{(j)} - B_{t_i^{(n)}}^{(j)})\}\{\sum_{j=1}^{k} g_{r'}^{(j)}(t_i^{(n)})(B_{t_{i+1}^{(n)}}^{(j)} - B_{t_i^{(n)}}^{(j)})\}\\
&\qquad\qquad \partial_r\partial_{r'}\varphi(t_i^{(n)}, \mathbf{Y}(t_i^{(n)}))\\
&\quad = \lim \sum_{j=1}^{k} g_r^{(j)}(t_i^{(n)})g_{r'}^{(j)}(t_i^{(n)}) \sum_{i=0}^{N_n-1} (B_{t_{i+1}^{(n)}}^{(j)} - B_{t_i^{(n)}}^{(j)})^2 \partial_r\partial_{r'}\varphi(t_i^{(n)}, \mathbf{Y}(t_i^{(n)}))\\
&\quad = \sum_{j=1}^{k} g_r^{(j)}(t_i^{(n)})g_{r'}^{(j)}(t_i^{(n)}) \lim \sum_{i=0}^{N_n-1} (t_{i+1}^{(n)} - t_i^{(n)})\partial_r\partial_{r'}\varphi(t_i^{(n)}, \mathbf{Y}(t_i^{(n)}))\\
&\quad = J_{12} := \sum_{j=1}^{k} \int_s^t g_r^{(j)}(u)g_{r'}^{(j)}(u)\partial_r\partial_{r'}\varphi(u, \mathbf{Y}(u))du.
\end{aligned} \tag{8.33}$$

The proof is completed by adding J_0, J_{11}, and J_{12}. ■

Applying Itô's lemma to the diffusion $\mathbf{Y}(t) = \mathbf{X}_t$ $(t \geq 0)$ constructed in Chapter 6, the following "SDE version", also called Itô's lemma, is obtained immediately.

Corollary 8.5 (Itô's Lemma for Diffusions) *Let $\mathbf{X}$ be a diffusion given by the solution to the stochastic differential equation (7.50), with $\alpha = 0$ and $\boldsymbol{\mu}(\cdot), \boldsymbol{\sigma}(\cdot)$ Lipschitzian. Assume $\varphi(t, \mathbf{y})$ satisfies the hypothesis of Theorem 8.4 with $m = k$.*

a. *Then one has the relation*

$$\varphi(t, \mathbf{X}_t) = \varphi(s, \mathbf{X}_s) + \int_s^t \{\partial_0\varphi(u, \mathbf{X}_u) + (\mathbf{A}\varphi)(u, \mathbf{X}_u)\}\, du$$
$$+ \sum_{r=1}^{k} \int_s^t \partial_r \varphi(u, \mathbf{X}_u)\boldsymbol{\sigma}_r(\mathbf{X}_u) \cdot d\mathbf{B}_u, \tag{8.34}$$

where $\mathbf{A}$ is the differential operator

$$(\mathbf{A}\varphi)(u, \mathbf{x}) := \frac{1}{2} \sum_{1 \leq r, r' \leq k} d_{rr'}(\mathbf{x})\partial_r\partial_{r'}\varphi(u, \mathbf{x}) + \sum_{r=1}^{k} \mu^{(r)}(\mathbf{x})\partial_r\varphi(u, \mathbf{x}), \tag{8.35}$$

for

$$((d_{rr'}(\mathbf{x}))) := \boldsymbol{\sigma}(\mathbf{x})\boldsymbol{\sigma}'(\mathbf{x}). \tag{8.36}$$

b. *In particular, if $\partial_0\varphi, \partial_{rr'}\varphi$ are bounded, then*

$$Z_t := \varphi(t, \mathbf{X}_t) - \int_0^t \{\partial_0\varphi(u, \mathbf{X}_u) + (\mathbf{A}\varphi)(u, \mathbf{X}_u)\}\, du \qquad (0 \leq t \leq T), \tag{8.37}$$

is a $\{\mathcal{F}_t\}$-martingale.

Note that part (b) is an immediate consequence of part (a), since $Z_t - Z_s$ equals the stochastic integral appearing in the right-hand side of (8.34), whose conditional expectation, given $\mathcal{F}_s$, is zero.

Remark 8.1 (Itô's Lemma for Diffusions with Locally Lipschitz Coefficients) An appropriate version of Corollary 8.5 extends to the case of diffusions with locally Lipschitz coefficients. For example, by Theorem 8.4, (8.34) holds when t is changed to $t \wedge \tau$ where $\tau = \inf\{u \geq 0 : |X_u^x| \geq R\}$ for every R such that $|x| < R$.

The (symmetric) matrix $\mathbf{H}_\varphi := ((\partial_r\partial_{r'}\varphi))_{1 \leq r, r' \leq m}$ appearing in the definition of $\mathbf{A}\varphi$ is referred to as the *Hessian matrix*. Borrowing other notions from vector calculus, such as the gradient vector of the (scalar) function φ defined by $\mathbf{grad}\,\varphi(t, \mathbf{x}) \equiv \nabla\varphi(t, \mathbf{x}) := (\frac{\partial\varphi}{\partial x_r}(t, \mathbf{x}))_{1 \leq r \leq m}$, one may more succinctly express the operator $\mathbf{A}$ as

$$\mathbf{A}\varphi(t,\mathbf{x}) = \frac{1}{2}\text{Trace}\mathbf{DH}(t,\mathbf{x}) + \boldsymbol{\mu}\cdot\nabla\varphi(t,\mathbf{x}). \tag{8.38}$$

The Hessian may also be recast as $H_\varphi = \nabla^t\nabla\varphi$.

As will be illustrated in later examples and exercises, a useful result for the product of two one-dimensional diffusions Z_1 and Z_2, say, may be obtained from the higher dimensional version of Itô's lemma as follows.

Proposition 8.6 *Suppose that* $\mathbf{Z} = (Z_1, Z_2)$ *is given by the stochastic differential equation*

$$d\mathbf{Z}(t) = \boldsymbol{\mu}(\mathbf{Z}(t))dt + \boldsymbol{\sigma}(\mathbf{Z}(t))d\mathbf{B}_t,\ \ \mathbf{Z}(0) = (x, y),$$

for Lipschitzian $\boldsymbol{\mu}, \boldsymbol{\sigma}$. *Then, letting* $\tilde{\mathbf{Z}} = (Z_2, Z_1)$, $D = \boldsymbol{\sigma}\boldsymbol{\sigma}^t = ((d_{r,r'}))_{1\le r,r'\le 2}$,

$$\begin{aligned} Z_1(t)Z_2(t) = z_1z_2 &+ \int_0^t \{\boldsymbol{\mu}(\mathbf{Z}(s))\cdot\tilde{\mathbf{Z}}(s) + d_{12}(\mathbf{Z}(s))\}ds \\ &+ \int_0^t \{Z_2(s)\sigma_{11}(\mathbf{Z}(s)) + Z_1(s)\sigma_{21}(\mathbf{Z}(s))\}dB_s^{(1)} \\ &+ \int_0^t \{Z_2(s)\sigma_{12}(\mathbf{Z}(s)) + Z_1(s)\sigma_{22}(\mathbf{Z}(s))\}dB_s^{(2)}, \end{aligned} \tag{8.39}$$

where

$$d_{12}(\mathbf{Z}(s)) = \sigma_{11}(\mathbf{Z}(s))\sigma_{21}(\mathbf{Z}(s)) + \sigma_{12}(\mathbf{Z}(s))\sigma_{22}(\mathbf{Z}(s)).$$

Proof Apply Itô's lemma to $\varphi(Z_1(t), Z_2(t)) = Z_1(t)Z_2(t)$ noting that $H_\varphi(z_1, z_2) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, and $\nabla\varphi(z_1, z_2) = (z_2, z_1)$. Along the way, also make use of the fact that $d_{12} = \sigma_{11}\sigma_{21} = d_{21}$ so that $\mathbf{D}H_\varphi(z_1, z_2) = \begin{pmatrix} d_{12} & d_{11} \\ d_{22} & d_{12} \end{pmatrix}$. In particular $\frac{1}{2}Trace(\mathbf{D}H_\varphi)(\mathbf{z}) = d_{12}$ and $\boldsymbol{\mu}\cdot\nabla\varphi(\mathbf{z}) = \boldsymbol{\mu}(\mathbf{z})\cdot\tilde{\mathbf{z}}$ for $\mathbf{z} = (z_1, z_2)$. ■

Remark 8.2 (Integration by Parts) Proposition 8.6 may be viewed as another version of "integration by parts". In particular, in the notation introduced there, and

$$dZ_i(t) = \mu_(\mathbf{Z}(t))dt + \sigma_i(\mathbf{Z}(t))dB_t, \quad i = 1, 2,$$

as a mnemonic device, one may formally express the result as

$$\int_0^t Z_1(s)dZ_2(s) = Z_1(s)Z_2(s)|_0^t - \int_0^t Z_2(s)dZ_1(s) - \int_0^t d_{12}(\mathbf{Z}(s))ds,$$

where $d_{12}(\mathbf{Z}) = \sigma_1(\mathbf{Z})\cdot\sigma_2(\mathbf{Z})$.

Example 3 (Langevin Equation and Ornstein–Uhlenbeck Process) Recall the Langevin equation defining the Ornstein–Uhlenbeck process,

$$dX(t) = -\gamma X(t)dt + \sigma dB_t, \quad X_0 = x_0, \tag{8.40}$$

which can be solved by Picard iteration (see Chapter 7, Example 1). Alternatively, by Itô's lemma, taking $\varphi(t, x) = e^{\gamma t}x$, one may check that

$$d(e^{\gamma t}X(t)) = \sigma e^{\gamma t}dB_t.$$

Thus,

$$X(t) = x_0 e^{-\gamma t} + \sigma \int_0^t e^{-\gamma(t-s)}dB_s.$$

Applying integration by parts (Proposition 6.6) To the stochastic integral, one obtains the previously derived formula

$$X(t) = x_0 e^{-\gamma t} - \gamma\sigma \int_0^t e^{-\gamma(t-s)}B_s ds + \sigma B_t.$$

Example 4 (Mean Reverting Ornstein–Uhlenbeck) Consider the process defined by

$$dX(t) = (\mu - \gamma X(t))dt + \sigma dB(t),\ X(0) = x,$$

where $\gamma > 0$. Applying Itô's lemma to $\varphi(t, X(t)) = e^{\gamma t}X(t)$, $\varphi(t, y) = e^{\gamma t}y$, one obtains the solution

$$X(t) = \frac{\mu}{\gamma} + (x - \frac{\mu}{\gamma})e^{-\gamma t} + \sigma \int_0^t e^{-\gamma(t-s)}dB(s), t \geq 0.$$

Example 5 (Geometric Brownian Motion) Consider the stochastic differential equation

$$dX(t) = \mu X(t)dt + \sigma X(t)dB_t,\ X(0) = x_0 > 0, \tag{8.41}$$

arising, for example, in mathematical finance as a model of "yield" $dX(t)/X(t)$ on a stock of price $X(t)$ at time t. One may not a priori conclude from the defining equation (8.41) that $\tau_0 = \inf\{t \geq 0 : X(t) = 0\} = +\infty$ a.s. However, notice that on the event $[\tau_0 > t]$, Itô's lemma provides that $d\log X(t) = \frac{1}{X(t)}dX(t) - \{\frac{1}{2}\sigma^2 X^2(t)/X^2(t)\}dt = \mu dt - \frac{1}{2}\sigma^2 dt + \sigma dB_t$. Thus,

$$\log X(t) = \log x_0 + (\mu - \frac{1}{2}\sigma^2)t + \sigma B_t, 0 \leq t < \tau_0.$$

Now observe that the *geometric Brownian motion* defined by

$$X(t) \equiv \varphi(t, B_t) = x_0 \exp\{\sigma B_t + \mu t - \frac{\sigma^2}{2} t\},$$

solves (8.41) for all times $t \geq 0$. In particular, it follows that $\tau_0 = \infty$ with probability one.

Example 6 (Affine Linear Coefficients) Consider the stochastic differential equation

$$dX(t) = (a_1 + b_1 X(t))dt + (a_2 + b_2 X(t))dB_t, t > 0, X(0) = x_0 > 0.$$

Writing

$$dX(t) = (b_1 X(t)dt + b_2 X(t))dB_t) + (a_1 dt + a_2 dB_t),$$

the solution is verified by the integration by parts formula applied to a method analogous to the integrating factor or, more generally, "variation of parameters"method of ordinary differential equations; see Exercise 10. Let Z_1 denote the geometric Brownian motion starting at one defined by

$$dZ_1(t) = b_1 Z_1(t)dt + b_2 Z_1(t)dB_t, t > 0, Z_1(0) = 1.$$

Also define

$$Z_2(t) = x_0 + \int_0^t (a_1 - a_2 b_2) Z_1^{-1}(s)ds + \int_0^t a_2 Z_1^{-1}(s)dB_s, \quad t \geq 0,$$

where

$$Z_1^{-1}(t) = \exp\{-b_2 B_t - (b_1 - \frac{b_2^2}{2})t\}, t \geq 0.$$

Now, using the integration by parts formula for the product given in Proposition 8.6, it follows that $X(t) = Z_1(t)Z_2(t)$. In particular,

$$\begin{aligned} X(t) = e^{\{b_2 B_t + (b_1 - \frac{b_2^2}{2})t\}} \{x_0 + \int_0^t (a_1 - a_2 b_2) e^{-b_2 B_s - (b_1 - \frac{b_2^2}{2})s} ds \\ + \int_0^t a_2 e^{-b_2 B_s - (b_1 - \frac{b_2^2}{2})s} dB_s\}. \end{aligned} \tag{8.42}$$

Example 7 (Logistic Population Growth Model) Consider the stochastic differential equation

$$dX(t) = rX(t)(k - X(t))dt + \sigma X(t)dB(t), \ X(0) = x > 0,$$

where r, k are nonnegative parameters describing the *growth rate* and *carrying capacity*, respectively. Observe that starting from $x > 0$, the process cannot reach zero. This follows from results in Section 8.1, especially (8.55). This is an example of an *inaccessible boundary* to be more fully discussed in Chapter 21. Consider $Y(t) = \frac{1}{X(t)}, t \geq 0$. From Itô's lemma, one obtains

$$dY(t) = rdt + (\sigma^2 - rk)Y(t)dt - \sigma Y(t)dB(t), \quad Y(0) = x^{-1}.$$

In particular, therefore, in the notation of the previous Example 6, Y is a process with affine linear coefficients $a_1 = r, b_1 = \sigma^2 - rk$ and $a_2 = 0, b_2 = \sigma$. The solution follows from (8.42). Specifically, one has

$$X(t) = \frac{1}{Y(t)} = \frac{k \exp\{\sigma B_t + (kr - \sigma^2/2)t\}}{x^{-1} + r\int_0^t \exp\{\sigma B_s + (kr - \sigma^2/2)s\}ds}, \quad t \geq 0. \tag{8.43}$$

Remark 8.3 Throughout this and subsequent chapters, unless otherwise specified, the stopping times are $\{\mathcal{F}_t : t \geq 0\}$–stopping times where $\{\mathcal{F}_t : t \geq 0\}$ is a P–complete filtration for which (6.9) holds (for multidimensional **B** when appropriate).

8.1 Asymptotic Properties of One-Dimensional Diffusions: Transience and Recurrence

While the focus of this subsection is on a complete characterization of the transience/recurrence dichotomy for one-dimensional diffusions, the following lemma obviously applies more generally.

To set some notation, denote the open ball of radius R and its boundary, respectively, by

$$B(\mathbf{x}_0 : R) := \{\mathbf{x} : |\mathbf{x} - \mathbf{x}_0| < R\}, \quad \partial B(\mathbf{x}_0 : R) = \{\mathbf{x} : |\mathbf{x} - \mathbf{x}_0| = R\}. \tag{8.44}$$

Also define the *escape time* from $B(\mathbf{x}_0 : R)$ by a process $\mathbf{X}^{\mathbf{x}} = \{\mathbf{X}_t^{\mathbf{x}} : t \geq 0\}$ by

$$\tau \equiv \tau_{\partial B}^{(\mathbf{x})} := \inf\{t \geq 0 : |\mathbf{X}_t^{\mathbf{x}} - \mathbf{x}_0| = R\}. \tag{8.45}$$

Lemma 1 *Let $\boldsymbol{\mu}(\mathbf{x})$, $\boldsymbol{\sigma}(\mathbf{x})$ be locally Lipschitzian on $\mathbb{R}^k$. Write $\boldsymbol{\sigma}_i(\mathbf{x})$ for the ith row of the matrix $\boldsymbol{\sigma}(\mathbf{x})$. Assume that, for some i, $d_{ii}(\mathbf{x}) = |\boldsymbol{\sigma}_i(\mathbf{x})|^2 > 0 \ \forall \ \mathbf{x} \in \mathbb{R}^k$. Then, for every $\mathbf{x} \in B(\mathbf{x}_0 : R)$, $\mathbb{E}\tau_{\partial B}^{(\mathbf{x})} < \infty$.*

Proof Without loss of generality, let $i = 1$. Let φ be a twice *continuously differentiable* function with compact support that equals $-\exp\{cx^{(1)}\}$ in $\overline{B(\mathbf{x}_0 : R)} := \{\mathbf{x} : |\mathbf{x} - \mathbf{x}_0| \le R\}$ for $c > 0$ ($\mathbf{x} = (\mathbf{x}^{(1)}, \ldots, x^{(k)})$). Then $A\varphi(\mathbf{x}) = (-\frac{1}{2}c^2 d_{11}(\mathbf{x}) - c\mu^{(1)}(\mathbf{x}))\exp\{cx^{(1)}\}$, so that for c large enough, there exists $\delta > 0$ satisfying $A\varphi(\mathbf{x}) \le -\delta < 0$ for $\mathbf{x} \in \overline{B(\mathbf{x}_0 : R)}$. By Itô's lemma (Corollary 8.5), $\varphi(\mathbf{X}_t) - \int_0^t A\varphi(\mathbf{X}_s)ds$ ($t \ge 0$) is a martingale whose expectation is $\varphi(\mathbf{x})$ for all $t \ge 0$, if $\mathbf{X}_0 \equiv \mathbf{x}$. Clearly, $\tau(\omega) \wedge t \le t < \infty$ for all $\omega \in \Omega$, and $|\mathbf{X}^{\mathbf{x}}_{\tau\wedge t}|$ is uniformly bounded and therefore uniformly integrable for all $t \ge 0$. Thus, by the optional stopping theorem (see BCPT[1] p. 61), for $\mathbf{X}_0 \equiv \mathbf{x} \in B(\mathbf{x}_0 : R) = \{\mathbf{z} : |\mathbf{z} - \mathbf{x}_0| < R\}$, one has

$$\mathbb{E}(\varphi(\mathbf{X}^{\mathbf{x}}_{\tau\wedge t})) - \mathbb{E}\int_0^{\tau\wedge t} A\varphi(\mathbf{X}^{\mathbf{x}}_s)ds = \varphi(\mathbf{x}) = -\exp\{cx^{(1)}\}.$$

But for $s \le \tau \wedge t$, $\mathbf{X}^{\mathbf{x}}_s \in \overline{B(\mathbf{x}_0 : R)}$, and the integral above is no more than $-\delta(\tau \wedge t)$, so that $-\mathbb{E}\int_0^{\tau\wedge t} A\varphi(\mathbf{X}^{\mathbf{x}}_s)ds \ge \delta\mathbb{E}(\tau \wedge t)$. Hence,

$$\delta\mathbb{E}(\tau \wedge t) \le \varphi(\mathbf{x}) - \mathbb{E}\varphi(\mathbf{X}^{\mathbf{x}}_{\tau\wedge t}) \le g(\mathbf{x}) := \varphi(\mathbf{x}) - \min\{\varphi(\mathbf{y}) : |\mathbf{y} - \mathbf{x}_0| \le R\} < \infty.$$

Now, let $t \uparrow \infty$ to get $\mathbb{E}\tau \le g(\mathbf{x})/\delta$. ∎

Let $X^x = \{X^x_t : t \ge 0\}$, $x \in \mathbb{R}$, be the diffusion on $\mathbb{R}$ with locally *Lipschitzian* coefficients $\mu(\cdot)$, $\sigma(\cdot)$ with $\sigma^2(x) > 0$ for all x. Let $\tau^x_a := \overline{\tau}_a \circ X^x$ be the first passage of X^x to $\{a\}$, $\tau^x_{a,b} := \tau^x_a \wedge \tau^x_b$. For a given pair $c < d$, define

$$\begin{aligned}\psi(x; c, d) &:= P(X^x \text{ reaches } c \text{ before } d), \qquad (8.46)\\ \varphi(x; c, d) &:= P(X^x \text{ reaches } d \text{ before } c).\end{aligned}$$

By Lemma 1, $P(\tau^x_{c,d} < \infty) \equiv P(X^x \text{ reaches } \{c, d\} \text{ in finite time}) = 1$, i.e., $\varphi(x; c, d) = 1 - \psi(x; c, d)$.

Proposition 8.7 *Under the above hypothesis on* $\mu(\cdot)$, $\sigma(\cdot)$,

a. $\varphi(x; c, d)$ *is the unique continuous solution to the equation*

$$A\varphi(x) = 0 \qquad \text{for } c < x < d, \quad \varphi(c) = 0, \varphi(d) = 1, \qquad (8.47)$$

where

$$A\varphi(x) := \frac{1}{2}\sigma^2(x)\varphi''(x) + \mu(x)\varphi'(x). \qquad (8.48)$$

[1] Throughout, BCPT refers to Bhattacharya and Waymire (2016), A Basic Course in Probability Theory.

b. *This solution is given by*

$$\varphi(x; c, d) = \int_c^x \exp\{-I(c, y)\}dy \Big/ \int_c^d \exp\{-I(c, y)\}dy \qquad (c < x < d),$$
$$I(c, y) := \int_c^y [2\mu(z)/\sigma^2(z)]dz. \tag{8.49}$$

Proof That the function φ in (8.49) satisfies (8.47) is easy to check. Also, one may easily extend φ to a twice continuously differentiable function on $\mathbb{R}$ with compact support (Exercise 4). Now, apply Itô's lemma (Corollary 8.5) to this extended function to get, using the optional sampling theorem

$$\mathbb{E}\varphi(X^x_{\tau^x_{c,d}\wedge t}) - \varphi(x) = \mathbb{E}\int_0^{\tau^x_{c,d}\wedge t} 0ds = 0. \tag{8.50}$$

Letting $t \uparrow \infty$, one obtains

$$\varphi(x) = \mathbb{E}\varphi(X^x_{\tau^x_{c,d}}) = P(X^x \text{ reaches } d \text{ before } c).$$

The uniqueness of the solution is easily shown, by directly solving (8.47) by integration (Exercise 5). ■

It is convenient to introduce notation for the (possibly infinite) *time of the last visit by* X^x as follows (with the convention that the supremum of the empty set is zero):

$$\eta^x_a := \sup\{t : X^x_t = a\}, \quad x, a \in \mathbb{R}. \tag{8.51}$$

Definition 8.2 A diffusion $\{X^x : x \in \mathbb{R}\}$ is said to be (point) *recurrent* if $\rho_{xa} := P(\tau^x_a < \infty) \equiv P(X^x \text{ reaches } a) = 1$ for all $x, a \in \mathbb{R}$. It is said to be *transient* if $P(\eta^x_a = \infty) = 0$ for all x, a.

Remark 8.4 It may be noted (Exercise 8) that $\{X^x, x \in \mathbb{R}\}$, is recurrent if and only if

$$P\left(\eta^x_a = \infty\right) = 1 \qquad \forall\, x, a \in \mathbb{R}, \tag{8.52}$$

recalling that the supremum of an empty set is taken to be zero in the definition of η^x_a.

Theorem 8.8 *Assume $\mu(\cdot), \sigma(\cdot)$ are locally Lipschitzian and that $\sigma(\cdot)$ does not vanish anywhere. Then (a) the diffusion is either recurrent, or it is transient. (b) The diffusion is recurrent if and only if*

$$\int_{-\infty}^{0} \exp\{I(y,0)\}dy = \infty \quad \textit{and} \quad \int_{0}^{\infty} \exp\{-I(0,y)\}dy = \infty. \tag{8.53}$$

Proof (b) Let $d > x$. Then, by (8.49),

$$\rho_{xd} \equiv P(\tau_d^x < \infty) = \lim_{c\downarrow-\infty} \varphi(x;c,d) = \lim_{c\downarrow-\infty} \frac{\int_c^x \exp\{-I(c,y)\}dy}{\int_c^d \exp\{-I(c,y)\}dy}.$$

Since $\exp\{-I(c,y)\} = \exp\{-I(c,0)\} \cdot \exp\{-I(0,y)\}$ (with the convention that $\int_a^b = -\int_b^a$ if $b < a$), one may cancel the common factor $\exp\{-I(c,0)\}$ from the numerator and the denominator to get

$$\begin{aligned}\rho_{xd} &= \lim_{c\downarrow-\infty} \frac{\int_c^x \exp\{-I(0,y)\}dy}{\int_c^d \exp\{-I(0,y)\}dy}.\\ &= \lim_{c\downarrow-\infty} \frac{\int_c^0 \exp\{-I(0,y)\}dy + \int_0^x \exp\{-I(0,y)\}}{\int_c^0 \exp\{-I(0,y)\}dy + \int_0^d \exp\{-I(0,y)\}dy}.\end{aligned}$$

The last limit is 1 if and only if $\lim_{c\downarrow-\infty} \int_c^0 \exp\{-I(0,y)\}dy = \infty$, which is the first relation in (8.53). Interpreting $\infty/\infty = 1$, one then has the general relation

$$\rho_{xd} = \frac{\int_{-\infty}^x \exp\{-I(0,y)\}dy}{\int_{-\infty}^d \exp\{-I(0,y)\}dy} \quad (x < d). \tag{8.54}$$

Similarly, if $c < x$, then letting $d \uparrow \infty$ In $\psi(x;c,d)$, one shows that

$$\rho_{xc} \equiv P(\tau_c^x < \infty) = \frac{\int_x^\infty \exp\{-I(0,y)\}dy}{\int_c^\infty \exp\{-I(0,y)\}dy}, \quad (c < x), \tag{8.55}$$

and it is 1 if and only if the second relation in (8.53) holds.

(a) Suppose that the process is not recurrent. Then there exist $c < d$ such that $\rho_{cd} < 1$ or $\rho_{dc} < 1$. We will first show that, for all $x \in (c,d)$, there is a last visit to c or d with probability one, i.e.,

$$P\left(\eta_c^x < \infty \quad \text{or} \quad \eta_d^x < \infty\right) = 1. \tag{8.56}$$

For this, define the stopping times

$$\begin{aligned}\tau_1 &:= \inf\{t \geq 0 : X_t^x = d\}, \qquad \tau_2 := \inf\{t > \tau_1 : X_t^x = c\},\\ \tau_{2r+1} &:= \inf\{t > \tau_{2r} : X_t^x = d\}, \quad \tau_{2r+2} := \inf\{t > \tau_{2r+1} : X_t^x = c\} \quad (r \geq 1).\end{aligned}$$

Then by the strong Markov property, writing $\theta_r := P(\tau_r < \infty)$, one has

$$\theta_1 = P(\tau_1 < \infty), \qquad \theta_{2r} = \theta_1(\rho_{dc}\rho_{cd})^{r-1}\rho_{dc}, \qquad \theta_{2r+1} = \theta_1(\rho_{dc}\rho_{cd})^r, \ r \geq 1.$$

Since $\rho_{dc}\rho_{cd} < 1$, $\sum_{r=1}^{\infty}\theta_r < \infty$. Therefore, (8.56) follows from the Borel–Cantelli lemma. Let us now use this to show that the process must be transient. Suppose it is not transient. Then, $\delta := P(\eta_a^x = \infty) > 0$ for some x, a. Define the stopping times

$$\gamma_1 := \inf\{t \geq 0 : X_t^x = a\}, \qquad \gamma_2 := \inf\{t > \gamma_1 + 1 : X_t^x = a\},$$
$$\gamma_{r+1} := \inf\{t > \gamma_r + 1 : X_t^x = a\} \qquad (r \geq 1).$$

Note that

$$P(\gamma_r < \infty) \geq \delta \text{ for every } r \geq 1,$$

which implies

$$P(\gamma_r < \infty \text{ and } X_{\gamma_r}^{+} \text{ reaches both} c \text{ and } d) \geq \delta\rho_{ac}\rho_{cd}\rho_{dc} > 0 \text{ for every } r \geq 1.$$

Since, by construction, $\gamma_{r+1} - \gamma_r \geq 1$, it follows that

$$P(\sup\{t : X_t^x = c\} > r \text{ and } \sup\{t : X_t^x = d\} > r) \geq \delta\rho_{ac}\rho_{cd}\rho_{dc} \text{ for all } r \geq 1,$$

so that

$$P(\eta_c^x = \infty \text{and } \eta_d^x = \infty) \geq \delta\rho_{ac}\rho_{cd}\rho_{cd} > 0.$$

This contradicts (8.56). ∎

Remark 8.5 Observe that in one-dimension and locally Lipschitz coefficients, the condition $\sigma^2(x) > 0$ for all $x \in \mathbb{R}$ implies that

$$P(X^x \text{will eventually reach } a) > 0 \quad \forall x, a \in \mathbb{R},$$

since (8.54) and (8.55) are always positive. Such a process is said to be *irreducible*.

Example 8 (Recurrence/Transience of One-Dimensional Brownian Motion) Consider $X_t^x = x + \mu t + \sigma B_t, t \geq 0$. Clearly, the criterion provides that the process is recurrent if and only if $\mu = 0$. Moreover, in this case, $P_x(\tau_d < \tau_c) = \frac{x-c}{d-c}$, $c \leq x \leq d$. More generally, see Exercise 9. For $\mu < 0$, say, one has for $c \leq x \leq d$,

$$P_x(\tau_d < \tau_c) = \frac{e^{\frac{2|\mu|}{\sigma^2}x} - e^{\frac{2|\mu|}{\sigma^2}c}}{e^{\frac{2|\mu|}{\sigma^2}d} - e^{\frac{2|\mu|}{\sigma^2}c}}. \tag{8.57}$$

In particular, letting $c \downarrow -\infty$, one has

$$P_x(\tau_d < \infty) = e^{-\frac{2|\mu|}{\sigma^2}(d-x)}. \tag{8.58}$$

Letting $M^x := \sup_{t\geq 0} X_t^x$, it follows that $M^x - x$ is exponentially distributed with parameter $\lambda = \frac{2|\mu|}{\sigma^2}$.

Exercises

1. (*Lévy martingale characterization of Brownian motion*) Show that if $X = \{X_t : t \geq 0\}$ is a continuous martingale starting at $x = 0$ with quadratic variation t, then X is standard Brownian motion.[*Hint:* Use Itô's lemma to check that for each $\lambda \in \mathbb{R}$, $\{e^{\lambda X_t - \frac{1}{2}\lambda^2 t} : t \geq 0\}$ is a martingale and deduce from the constant expected value property and moment generating function that $X_t - X_s$ is Gaussian with mean zero and variance $t - s$. Extend this calculation to increments $X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_m} - X_{t_m - 1}, 0 < t_0 < t_1 < \cdots < t_m, m \geq 1$ by iteratively conditioning.]
2. (*Geometric Brownian Motion*) In Example 5, show that starting from $x > 0$, (a) $\{X_t : t \geq 0\}$ reaches every state $y > 0$ with positive probability, whatever be μ and $\sigma^2 > 0$, and (b) it reaches every state with probability one if $\mu \leq \frac{1}{2}\sigma^2$. (c) Compute the time-asymptotic behavior of X_t in each of the cases (i) $\mu > \frac{1}{2}\sigma^2$, (ii) $\mu < \frac{1}{2}\sigma^2$, and (iii) $\mu = \frac{1}{2}\sigma^2$. [*Hint:* Express X explicitly in terms of B.]
3. Give a proof of the integration by parts formula Proposition 6.6 using Itô's lemma.
4. Let $\varphi : [c, d] \to \mathbb{R}$ be twice continuously differentiable on $[c, d]$, with $\varphi'(c), \varphi''(c)$ denoting right-hand derivatives, and $\varphi'(d), \varphi''(d)$ denoting left-hand derivatives. Show that φ has a twice continuously differentiable extension with compact support to all of $\mathbb{R}$.
5. Prove uniqueness of the solution to (8.47). [*Hint:* Give an argument from which the solution is explicitly computed.]
6. (*Ornstein–Uhlenbeck*) Consider the Ornstein–Uhlenbeck process X defined by Example (3). Use Itô's lemma to compute (a) the mean and variance of X_t, and (b) the characteristic function $\mathbb{E}e^{irX_t} = \mathbb{E}\cos(rX_t) + i\mathbb{E}\sin(rX_t), r \in \mathbb{R}$. (c) Show that $\limsup_{t\to\infty} X_t = +\infty, \liminf_{t\to\infty} X_t = -\infty$ regardless of how small the parameter $\sigma \neq 0$ might be.
7. (*Multidimensional Ornstein–Uhlenbeck*) Let Γ be a $k \times k$ real matrix with rows $\boldsymbol{\gamma}_i, i = 1, \ldots, k$ and $\sigma > 0$. Let $\mathbf{B}$ denote $k-$dimensional standard Brownian motion and let $\mathbf{Y}$ denote the process defined by $dY^{(r)}(t) = \boldsymbol{\gamma}_r \cdot \mathbf{Y}(t)dt + \sigma dB_t^{(r)}, t > 0, Y^{(r)}(0) = y^{(r)}, r = 1, \ldots, k$. Equivalently, $d\mathbf{Y}(t) = \Gamma\mathbf{Y}(t)dt + \sigma d\mathbf{B}_t, \quad t > 0, \mathbf{Y}(0) = \mathbf{y}$. Defining $e^{\Gamma t} = \sum_{n=0}^{\infty} \frac{\Gamma^n t^n}{n!}$, note that Γ commutes with $e^{\Gamma t}$ and $\frac{d}{dt}e^{\Gamma t} = \Gamma e^{\Gamma t}$. Establish the representation

has (component-wise) $\mathbf{Y}(t) = e^{\Gamma t}\mathbf{y} + \mathbf{B}_t - \sigma \int_0^t \Gamma e^{\Gamma(t-s)}\mathbf{B}_s ds, t \geq 0$.[*Hint:* Denote the ij-entry of $e^{\Gamma t}$ by $\varepsilon_{ij}(t)$and check that $\frac{d}{dt}\varepsilon_{ij}(t) = \sum_{m=1}^{k} \varepsilon_{im}(t)\gamma_{mj}$. Define $\varphi_r(t, \mathbf{y}) = \sum_{j=1}^{k} \varepsilon_{rj}(t)y^{(j)}$ and use Itô's lemma to show $d\varphi_r(t, \mathbf{Y}(t)) = \sigma_{j=1}^{k}\varepsilon_{rj}(t)dB_t^{(j)}$.]

8. Show that the condition (8.52) is necessary and sufficient for (point) recurrence as defined in Definition 8.2.
9. Show that a one-dimensional diffusion with zero drift and strictly positive Lipschitz diffusion coefficient is always a recurrent process.
10. (*Variation of Parameters*) Consider the ordinary differential equation $x'(t) = a(t)x(t) + b(t), t > 0, x(0) = x_0$. Assume $\psi(t) = e^{\int_0^t a(s)ds}$ is a well-defined, positive solution to $\psi'(t) = a(t)\psi(t), t \geq 0$. Show that $x(t) = \psi^{-1}(t)(x_0 + \int_0^t b(s)\psi^{-1}(s)ds), t \geq)$ is a solution to the given problem.

Chapter 9
Cameron–Martin–Girsanov Theorem

The topic of this chapter is a useful change-of-measure formula based on the mutual absolute continuity, and a relatively simple formula for the Radon-Nikodym derivative, of the distributions of a pair of diffusions over a finite time interval that have the same diffusion coefficient. Two derivations are provided of this important technique, each with its own merits.

A *change-of-measure formula* under a drift change for nonsingular diffusions is one of the cornerstones of the theory of stochastic differential equations. To have some feel for the theory, consider two Brownian motions, say standard Brownian motion $B = \{B_t : t \geq 0\}$ starting at zero and Brownian motion $X = \{X_t \equiv B_t + \mu t : t \geq 0\}$ starting at zero having nonzero drift μ. The distributions of B and X are mutually singular probabilities on the path space $C[0, \infty)$, e.g. by the strong law of large numbers (see Exercise 11). However, restricted to a finite time horizon $0 \leq t \leq T < \infty$, the distributions of $B_{[0,T]} := \{B_t : 0 \leq t \leq T\}$ and $X_{[0,T]} := \{X_t : 0 \leq t \leq T\}$ will be seen to be mutually absolutely continuous on $C[0, T]$. This and much more, including a formula for the Radon-Nikodym derivative, make up the Cameron–Martin–Girsanov theory for diffusions. In general, the Radon-Nikodym derivative provides a useful transformation between probability distributions of diffusions with different drifts but common nonsingular diffusion coefficients. The change-of-measure formula was first derived by Cameron and Martin (1944), (1945), for the computation of distributions of functionals of Brownian motion under translation by nonrandom as well as random functions. The present generalization to all nonsingular diffusions is due to Girsanov (1960). We provide two different approaches toward its derivation, each with its own merits. Basic to both approaches is the following result on *exponential martingales*.

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_9

As usual, let $(\Omega, \mathcal{F}, P)$ be a probability space on which is defined a k-dimensional standard Brownian motion $\mathbf{B}$, with respect to a P-complete filtration $\{\mathcal{F}_t : t \geq 0\}$.

Proposition 9.1 *Let* $\mathbf{f}_t$, $\alpha \leq t \leq T$, *be a nonanticipative bounded functional with values in* $\mathbb{R}^k$. *Then*

$$M_t := \exp\{\int_\alpha^t \mathbf{f}_s \cdot d\mathbf{B}_s - \frac{1}{2}\int_\alpha^t |\mathbf{f}_s|^2 ds\}, \quad \alpha \leq t \leq T, \tag{9.1}$$

is a $\{\mathcal{F}_t : t \geq 0\}$*-martingale. In particular,* $EM_t = EM_\alpha = 1$, $\alpha \leq t \leq T$.

Proof First assume $\mathbf{f}$ is a bounded nonanticipative step functional: $\mathbf{f}_t = \mathbf{g}_i$ for $t_i \leq t < t_{i+1}$ ($\alpha = t_0 < t_1 < \cdots < t_m = T$), $0 \leq i \leq m-1$, $\mathbf{f}_T = \mathbf{g}_m$, where $\mathbf{g}_i$ is $\mathcal{F}_{t_i}$-measurable and bounded ($0 \leq i \leq m$). If $t \in [t_j, t_{j+1})$, and $s \in [t_i, t_{i+1})$, $i < j$, then $\mathcal{F}_s \subseteq \mathcal{F}_{t_j}$ so that

$$\begin{aligned}\mathbb{E}[M_t \mid \mathcal{F}_s] &= \mathbb{E}[\mathbb{E}(M_t \mid \mathcal{F}_{t_j}) \mid \mathcal{F}_s] \\ &= \mathbb{E}[M_{t_j}\mathbb{E}(\exp\{\mathbf{g}_j \cdot (\mathbf{B}_t - \mathbf{B}_{t_j}) - \frac{1}{2}(t-t_j)|\mathbf{g}_j|^2\} \mid \mathcal{F}_{t_j})|\mathcal{F}_s] \\ &= \mathbb{E}[M_{t_j} \mid \mathcal{F}_s],\end{aligned} \tag{9.2}$$

since the conditional distribution of $\mathbf{g}_j \cdot (\mathbf{B}_t - \mathbf{B}_{t_j})$, given $\mathcal{F}_{t_j}$, is normal with mean zero and variance $(t-t_j)|\mathbf{g}_j|^2$, and $\mathbb{E}\exp\{Y - \frac{\sigma^2}{2}\} = 1$ for every normal random variable Y with mean zero and variance σ^2 (i.e., $\theta \to \exp\{\theta^2\sigma^2/2\}$ is the moment generating function of Y). If $i < j-1$, then $\mathcal{F}_s \subseteq \mathcal{F}_{t_{j-1}}$ and taking conditional expectation of M_{t_j} given $\mathcal{F}_{t_{j-1}}$, as in (9.2), one gets

$$\mathbb{E}[M_t \mid \mathcal{F}_s] = \mathbb{E}[M_{t_j} \mid \mathcal{F}_s] = \mathbb{E}[M_{t_{j-1}} \mid \mathcal{F}_s].$$

Proceeding in this manner, one finally arrives at

$$\begin{aligned}\mathbb{E}[M_t \mid \mathcal{F}_s] &= \mathbb{E}[M_{t_{i+1}} \mid \mathcal{F}_s] \\ &= M_s\mathbb{E}(\exp\{\mathbf{g}_i \cdot (\mathbf{B}_{t_{i+1}} - \mathbf{B}_s) - \frac{1}{2}(t_{i+1}-s)|\mathbf{g}_i|^2\} \mid \mathcal{F}_s) \\ &= M_s.\end{aligned}$$

If $s, t \in [t_j, t_{j+1})$, $s < t$, then $M_t = M_s \exp\{\mathbf{g}_j(\mathbf{B}_t - \mathbf{B}_s) - \frac{1}{2}(t-s)|\mathbf{g}_j|^2\}$, and $\mathbb{E}[M_t \mid \mathcal{F}_s] = M_s \cdot 1 = M_s$, proving the martingale property.

Now let $\mathbf{f}$ be an arbitrary bounded nonanticipative step functional on $[\alpha, T]$. There exists a uniformly bounded (by a constant A, say) sequence of nonanticipative step functionals $\mathbf{f}^{(n)}$ such that $Y_t^{(n)} := \int_\alpha^t \mathbf{f}_s^{(n)} \cdot d\mathbf{B}_s - \frac{1}{2}\int_\alpha^t |\mathbf{f}_s^{(n)}|^2 ds$ converges in probability to $\int_\alpha^t \mathbf{f}_s \cdot d\mathbf{B}_s - \frac{1}{2}\int_\alpha^t |\mathbf{f}_s|^2 ds = Y_t$, as shown in the proof of Proposition 6.3. Then, writing

$$M_t^{(n)} := \exp\{Y_t^{(n)}\}, \quad M_t := \exp\{Y_t\},$$

one has

$$\begin{aligned}\mathbb{E}(M_t^{(n)})^2 &= \mathbb{E}\exp\{2\int_\alpha^t \mathbf{f}_s^{(n)} \cdot d\mathbf{B}_s - 2\int_\alpha^t |\mathbf{f}_s^{(n)}|^2 ds\} \cdot \exp\{\int_\alpha^t |\mathbf{f}_s^{(n)}|^2 ds\} \\ &\leq 1 \cdot \exp\{(t-\alpha)A^2\} < \infty, \qquad (\alpha \leq t \leq T).\end{aligned}$$

Hence, $\{M_t^{(n)} : n \geq 1\}$ is uniformly integrable, and it converges to M_t in L^1 (as $n \to \infty$), $\alpha \leq t \leq T$. Therefore, taking limits in the identity $M_s^{(n)} = \mathbb{E}[M_t^{(n)} \mid \mathcal{F}_s]$ $(\alpha \leq s < t \leq T)$, one obtains the desired result: $M_s = \mathbb{E}[M_t \mid \mathcal{F}_s]$. ■

Corollary 9.2 *Let* $\mathbf{f} = \{\mathbf{f}_t : \alpha \leq t \leq T\} \in \mathcal{L}[\alpha, T]$, *i.e.,* $\mathbf{f}$ *is nonanticipative on* $[\alpha, T]$ *and* $\int_\alpha^T |\mathbf{f}_s|^2 ds < \infty$. *Then* $M_t := \exp\{\int_\alpha^t \mathbf{f}_s \cdot d\mathbf{B}_s - \frac{1}{2}\int_\alpha^t |\mathbf{f}_s|^2 ds\}$, $0 \leq t \leq T$, *is a supermartingale, i.e.,* $\mathbb{E}[M_t \mid \mathcal{F}_s] \leq M_s$. *In particular,* $\mathbb{E}M_t \leq 1$, $(\alpha \leq t \leq T)$.

Proof One may find a sequence of bounded nonanticipative functionals $\mathbf{f}^{(n)} = \{\mathbf{f}_t^{(n)} : \alpha \leq t \leq T\}$ that converge in probability and in $L^2(dt \times dP)$ to $\mathbf{f}$ as $n \to \infty$. Then $M_{s,t}^{(n)} := \exp\{\int_s^t \mathbf{f}_u^{(n)} \cdot d\mathbf{B}_u - \frac{1}{2}\int_s^t |\mathbf{f}_u^{(n)}|^2 du\} \to M_{s,t} := \exp\{\int_s^t \mathbf{f}_u \cdot d\mathbf{B}_u - \frac{1}{2}\int_s^t |\mathbf{f}_u|^2 du\}$ in probability. Hence, by Fatou's lemma, writing $M_t := M_{\alpha,t}$, $M_t^{(n)} := M_{\alpha,t}^{(n)}$, one has for all $A \in \mathcal{F}_s$ $(\alpha \leq s < t \leq T)$,

$$\begin{aligned}\mathbb{E}(M_t \mathbf{1}_A) &\equiv \mathbb{E}(M_s \mathbf{1}_A M_{s,t}) = \mathbb{E}(\lim_{n\to\infty} M_s \mathbf{1}_A M_{s,t}^{(n)}) \\ &\leq \lim_{n\to\infty} \mathbb{E}(M_s \mathbf{1}_A M_{s,t}^{(n)}) = \lim_{n\to\infty} \mathbb{E}[\mathbb{E}(M_s \mathbf{1}_A M_{s,t}^{(n)} \mid \mathcal{F}_s)] \\ &= \lim_{n\to\infty} \mathbb{E}[M_s \mathbf{1}_A \mathbb{E}(M_{s,t}^{(n)} \mid \mathcal{F}_s)] = \mathbb{E}(M_s \mathbf{1}_A),\end{aligned}$$

since $\mathbb{E}(M_{s,t}^{(n)} \mid \mathcal{F}_s) = 1$ by Proposition 9.1. ■

Corollary 9.3 *Suppose* $\mathbf{f} = (f^{(1)}, \ldots, f^{(k)})$ *is a nonanticipative functional on* $[0, T]$ *such that*

$$\mathbb{E}\exp\{\theta \int_0^T |\mathbf{f}(s)|^2 ds\} < \infty \tag{9.3}$$

for some $\theta > 1$. *Then* M_t, $0 \leq t \leq T$, *defined by (9.1) is a martingale.*

Proof Define $f_n^{(j)}(t) := (f^{(j)}(t) \wedge n) \vee (-n)$ $(1 \leq j \leq k)$. Then $\mathbf{f}_n = (f_n^{(1)}, \ldots, f_n^{(k)})$ is a bounded nonanticipative functional on $[0, T]$. Hence $M_t^{(n)} := \exp\{\int_0^t \mathbf{f}_n(s) \cdot d\mathbf{B}_s - \frac{1}{2}\int_0^t |\mathbf{f}_n(s)|^2 ds\}$, $0 \leq t \leq T$, is a martingale, by Proposition 9.1. Since $M_t^{(n)} \to M_t$ a.s. as $n \to \infty$, it is enough to prove that $M_t^{(n)}$, $n \geq 1$, is uniformly integrable. For $\theta_1 > 1$, one has

$$
\begin{aligned}
\mathbb{E}(M_t^{(n)})^{\theta_1} &= \mathbb{E}\exp\{\theta_1 \int_0^t \mathbf{f}_n(s)\cdot d\mathbf{B}_s - \frac{\theta_1}{2}\int_0^t |\mathbf{f}_n(s)|^2 ds\} \\
&= \mathbb{E}[\exp\{\theta_1 \int_0^t \mathbf{f}_n(s)\cdot d\mathbf{B}_s - \frac{\theta_1^3}{2}\int_0^t |\mathbf{f}_n(s)|^2 ds\} \\
&\quad \cdot \exp\{\frac{(\theta_1^3-\theta_1)}{2}\int_0^t |\mathbf{f}_n(s)|^2 ds\}].
\end{aligned}
$$

By Hölder's inequality, the last expectation is no greater than

$$
\begin{aligned}
&(\mathbb{E}\exp\{\theta_1^2 \int_0^t \mathbf{f}_n(s)\cdot d\mathbf{B}_s - \frac{\theta_1^4}{2}\int_0^t |\mathbf{f}_n(s)|^2 ds\})^{\frac{1}{\theta_1}} \\
&\cdot(\mathbb{E}\exp\{(\frac{\theta_1}{\theta_1-1})(\frac{\theta_1^3-\theta_1}{2})\int_0^t |\mathbf{f}_n(s)|^2 ds\})^{1-\frac{1}{\theta_1}} \\
= {}& 1\cdot(\mathbb{E}\exp\{\theta_1^2\frac{(\theta_1+1)}{2}\int_0^t |\mathbf{f}_n(s)|^2 ds\})^{1-\frac{1}{\theta_1}}.
\end{aligned}
$$

Choose θ_1 close enough to 1 such that $\frac{\theta_1^2(\theta_1+1)}{2} \le \theta$. Then

$$
\begin{aligned}
\mathbb{E}(M_t^{(n)})^{\theta_1} &\le (\mathbb{E}\exp\{\theta \int_0^t |\mathbf{f}_n(s)|^2 ds\})^{1-\frac{1}{\theta_1}} \\
&\le (\mathbb{E}\exp\{\theta \int_0^t |\mathbf{f}(s)|^2 ds\})^{1-\frac{1}{\theta_1}}.
\end{aligned}
$$

Hence, by (9.3), $\sup\{\mathbb{E}(M_t^{(n)})^{\theta_1} : n \ge 1\} < \infty$, so that $M_t^{(n)}$ is uniformly integrable and $M_t = \lim M_t^{(n)}$ is a martingale. ■

Remark 9.1 It has been proved by Novikov (1972) that if

$$
\mathbb{E}\exp\{\frac{1}{2}\int_0^T |\mathbf{f}(s)|^2 ds\} < \infty, \tag{9.4}
$$

i.e. $\theta = 1/2$ in (9.3), then M_t, $0 \le t \le T$, is a martingale.

For our first approach to the main result (CMG Theorem), we will need Paul Lévy's important characterization of Brownian motion as a continuous martingale with the appropriate quadratic variation.

Theorem 9.4 (Lévy's Martingale Characterization of Brownian Motion) *Let* $\mathbf{Z}_t = (Z_t^{(1)}, \ldots, Z_t^{(k)})$ $t \ge 0$, *be an* $\mathbb{R}^k$*-valued stochastic process adapted to a filtration* $\{\mathcal{F}_t : t \ge 0\}$ *of* $\mathcal{F}$, *defined on a probability space* $(\Omega, \mathcal{F}, P)$. *Then* $\mathbf{Z}_t - \mathbf{Z}_0$, *is a k-dimensional standard Brownian motion if and only if*

i. *$t \to \mathbf{Z}_t$ is continuous on* $[0, \infty)$*, a.s.,*
ii. *$Z_t^{(i)} - Z_0^{(i)}$, $t \geq 0$, is a $\{\mathcal{F}_t : t \geq 0\}$–martingale for each $i = 1, 2, \dots, k$, and*
iii. *$(Z_t^{(i)} - Z_0^{(i)})(Z_t^{(j)} - Z_0^{(j)}) - \delta_{ij}t$, $t \geq 0$, is a $\{\mathcal{F}_t : t \geq 0\}$-martingale for each pair (i, j), $1 \leq i \leq j \leq k$.*

Proof The "only if" part is obvious. To prove the "if" part, we need to show that for arbitrary m and $0 = t_0 < t_1 < \cdots < t_m$, $\mathbf{Z}_{t_j} - \mathbf{Z}_{t_{j-1}}$, $1 \leq j \leq m$, are independent Gaussian random vectors with zero mean vectors and dispersion matrices $(t_j - t_{j-1})I$, $1 \leq j \leq m$, where I is the $k \times k$ identity matrix. For this, it is enough to prove, for all such m and $t_0, t_1, \dots, t_m$, $i \leq j \leq m$,

$$\mathbb{E}\exp\{i\sum_{j=1}^{m}\boldsymbol{\xi}_j \cdot (\mathbf{Z}_{t_j} - \mathbf{Z}_{t_{j-1}})\} = \exp\{-\frac{1}{2}\sum_{j=1}^{m}(t_j - t_{j-1})|\boldsymbol{\xi}_j|^2\}, \boldsymbol{\xi}_j \in \mathbb{R}^k. \tag{9.5}$$

Without loss of generality, assume $\boldsymbol{\xi}^{(j)} \neq 0 \ \forall\, j$ (else, just delete those $\mathbf{Z}_{t_j} - \mathbf{Z}_{t_{j-1}}$ terms for which $\boldsymbol{\xi}^{(j)} = 0$). Write

$$\begin{aligned}
&t_{j-1,r,n} = t_{j-1} + r2^{-n}(t_j - t_{j-1}) \quad (r = 0, 1, \dots, 2^n),\\
&Z_{j,r,n} = \boldsymbol{\xi}_j \cdot (\mathbf{Z}_{t_{j-1,r,n}} - \mathbf{Z}_{t_{j-1,r-1,n}}) \quad (r = 1, \dots, 2^n),\\
&\sum_{j=1}^{m}\boldsymbol{\xi}_j \cdot (\mathbf{Z}_{t_j} - \mathbf{Z}_{t_{j-1}}) = \sum_{j=1}^{m}\sum_{r=1}^{2^n} Z_{j,r,n}.
\end{aligned} \tag{9.6}$$

Then, for each n, $Z_{j,r,n}$ $(1 \leq j \leq m, 1 \leq r \leq 2^n)$ are martingale differences when the indices (j, r) are ordered by the values of $t_{j-1,r,n} \equiv t_{j-1} + r2^{-n}(t_j - t_{j-1})$. One has, by (ii) and (iii),

$$\begin{aligned}
&\mathbb{E}(Z_{j,r,n} \mid \mathcal{F}_{t_{j-1,r,n-1}}) = 0,\\
&\sigma_{j,r,n}^2 := \mathbb{E}(Z_{j,r,n}^2 \mid \mathcal{F}_{t_{j-1,r,n-1}}) = 2^{-n}(t_j - t_{j-1})|\boldsymbol{\xi}_j|^2,\\
&s_n^2 := \sum_{j=1}^{m}\sum_{r=1}^{2^n}\sigma_{j,r,n}^2 = \sum_{j=1}^{m}(t_j - t_{j-1})|\boldsymbol{\xi}_j|^2,\\
&L_n(\varepsilon) := \sum_{j=1}^{m}\sum_{r=1}^{2^n}\mathbb{E}(Z_{j,r,n}^2\mathbf{1}_{[|Z_{j,r,n}|>\varepsilon]} \mid \mathcal{F}_{t_{j-1,r,n-1}}), \ (\varepsilon > 0).
\end{aligned} \tag{9.7}$$

By the martingale central limit theorem,[1] if we prove that $L_n(\varepsilon) \to 0$ in probability as $n \to \infty$, then (9.5) will follow from (9.7). For this, first note that $L_n(\varepsilon)$

[1] See Bhattacharya and Waymire (2022), Theorem 15.1.

is uniformly bounded (for all n and ε) by $\sum_{j=1}^{m}(t_j - t_{j-1})|\boldsymbol{\xi}_j|^2$. Also, by the assumption (i) of continuity, $\delta_n := \max\{|Z_{j,r,n}| : 1 \leq j \leq m, 1 \leq r \leq 2^n\} \to 0$ a.s. as $n \to \infty$. Therefore, $EL_n(\varepsilon) \to 0$ as $n \to \infty$, for every $\varepsilon > 0$. ■

Remark 9.2 The assumptions (ii), (iii) of Theorem 9.4 are satisfied by every square integrable process X_t, $t \geq 0$, with (homogeneous) independent mean-zero increments (if the process is scaled so as to have variance one for $X_{t+1} - X_t$), for the special case $k = 1$. For example, if N_t, $t \geq 0$, is a Poisson process with intensity parameter (mean per unit time) $\lambda = 1$, then $X_t := N_t - t$ satisfies (ii), (iii) with $k = 1$. Thus, the continuity assumption (i) is crucial for the martingale characterization of Brownian motion.

For the case of a stochastic integral $Z_t := \int_0^t \mathbf{f}(u) \cdot d\mathbf{B}_u$, a more direct proof of Lévy's characterization is sketched in Exercise 5.

As an important consequence of Lévy's characterization of Brownian motion, we get the following result that says that every stochastic integral is a standard Brownian motion when time is measured by quadratic variation. In other words, a large class of martingales (adapted to Brownian motion) with continuous paths can be expressed as time-changed Brownian motions.[2]

Corollary 9.5 (Time Change of Stochastic Integrals) *Let* $\mathbf{f} \in \mathcal{L}[0, \infty)$ *be such that* $t \to \int_0^t |\mathbf{f}(s)|^2 ds$ *is strictly increasing and goes to* ∞ *(a.s.) as* $t \to \infty$. *Define the stopping times* $\tau_t := \inf\{s \geq 0 : \int_0^s |\mathbf{f}(u)|^2 du = t\}$, $t \geq 0$. *Then* $Y_t := I_{\mathbf{f}}(\tau_t)$, $t \geq 0$, *is a standard one-dimensional Brownian motion, where* $I_{\mathbf{f}}(s) := \int_0^s \mathbf{f}(u) \cdot d\mathbf{B}_u$ $(s \geq 0)$.

Proof Since the inverse of a strictly increasing continuous function on $[0, \infty)$ into $[0, \infty)$ is continuous, $t \to \tau_t$ is continuous (a.s.) and so is $t \to Y_t$. Now, assuming first that $\mathbf{f} \in \mathcal{M}[0, \infty)$, $I_{\mathbf{f}}(t)$, $t \geq 0$, is a $\{\mathcal{F}_t : t \geq 0\}$–martingale. Hence, by the optional stopping rule, $I_{\mathbf{f}}(\tau_t)$, $t \geq 0$, is a $\{\mathcal{F}_{\tau_t} : t \geq 0\}$–martingale. Also, since $S_t := I_{\mathbf{f}}^2(t) - \int_0^t |\mathbf{f}(s)|^2 ds$, $t \geq 0$, is a martingale, so is $X_t := S_{\tau_t}$ $(t \geq 0)$. But $X_t = Y_t^2 - t, t \geq 0$. Hence, by Lévy's characterization, $Y_t, t \geq 0$, is a standard one-dimensional Brownian motion. For a general $\mathbf{f} \in \mathcal{L}[0, \infty]$, let $\mathbf{f}_n(t) := \mathbf{f}(t)\mathbf{1}_{[\tau_n > t]}$ $(n = 1, 2, \ldots)$. Then $\mathbb{E} \int_0^T |\mathbf{f}_n(u)|^2 du \leq n \,\forall\, T$, so that $\mathbf{f}_n \in \mathcal{M}[0, \infty]$. Therefore, a minor modification of Theorem 9.4 implies $Y_t^{(n)} := I_{\mathbf{f}_n}(\tau_t^{(n)})$ is a standard Brownian motion on $[0, n]$, where $\tau_t^{(n)} := \inf\{s \geq 0 : \int_0^s |\mathbf{f}(u)|^2 \mathbf{1}_{[\tau_n > u]} du = t\} = \inf\{s \geq 0 : \int_0^{s \wedge \tau_n} |\mathbf{f}(u)|^2 du = t\} = \tau_t$ for $t \leq n$ (since $\tau_n \geq \tau_t$). Thus, $Y_t^{(n)} = I_{\mathbf{f}_n}(\tau_t) = I_{\mathbf{f}}(\tau_t)$, $0 \leq t \leq n$. Hence, $I_{\mathbf{f}}(\tau_t)$ is a standard Brownian motion on $[0, n]$ for every $n \geq 1$. ■

[2] In a startling and sensational recent discovery of a sealed letter written some 60 years ago to the Academy of Sciences in Paris, it was revealed that by the age of 25, Wolfgang Doeblin had obtained this result among several others in the development of the stochastic calculus. See Handwerk and Willems (2007), Davis and Etheridge (2006) for historical accounts.

Remark 9.3 The assumption $t \to \int_0^t |\mathbf{f}(u)|^2 du$ is strictly increasing may be dispensed by defining $\tau_t := \inf\{s \geq 0 : \int_0^s |\mathbf{f}(u)|^2 du > t\}$. τ_t is then an $\{\mathcal{F}_t : t \geq 0\}$–optional time, which is also a $\{\mathcal{F}_t\}$–stopping time if $\{\mathcal{F}_t : t \geq 0\}$ is right continuous (see BCPT[3] p. 59). One may thus take the filtration to be $\mathcal{F}_{t+} := \cap_{\varepsilon>0}\mathcal{F}_{t+\varepsilon}$ $(t \geq 0)$ if necessary. It may be shown that, outside a P–null set, $t \to I_{\mathbf{f}}(\tau_t)$ is continuous.[4] The proof of the desired martingale properties of $Y_t := I_{\mathbf{f}}(\tau_t)$ and $Y_t^2 - t$ $(t \geq 0)$ follows as above.

Example 1 (One-Dimensional Diffusion) Consider a one-dimensional diffusion satisfying

$$dX(t) = \sigma(X(t))dB(t), \quad X(0) = x,$$

where σ is a positive Lipschitz function on $\mathbb{R}$ bounded away from zero. To "find"the time-change τ_t that makes $Z(t) = X(\tau_t), t \geq 0$, a standard Brownian motion, consider that a continuous strictly increasing process $t \to \tau_t$, $\tau_0 = 0$, would make $Z(t), t \geq 0$ a continuous martingale. Thus, in view of Corollary 9.5, it is sufficient to determine such a process $\tau_t, t \geq 0$, for which Z has the same quadratic variation as Brownian motion, namely, t. Now, by Itô's lemma,

$$Z^2(t) - t = X^2(\tau_t) - t = x^2 + \int_0^{\tau_t} \sigma^2(X_s)ds - t + \int_0^t \sigma(X_s)dB_s, t \geq 0.$$

So the martingale property of $Z^2(t) - t, t \geq 0$, may be cast as a definition of τ_t as follows

$$\int_0^{\tau_t} \sigma^2(X(s))ds = t, \quad t \geq 0.$$

Indeed, it follows from Corollary 9.5 that $Z(t) = X(\tau_t), t \geq 0$, is standard Brownian motion starting at x. Conversely, $X(t) = B(\tau_t^{-1}), t \geq 0$, can be viewed as a time change of Brownian motion, where τ_t^{-1} is given by by

$$\int_0^{\tau_t^{-1}} \frac{ds}{\sigma^2(X(s))} = t, \quad t \geq 0.$$

Remark 9.4 One may relax the assumption on $\sigma(\cdot)$ to be just measurable and bounded away from zero, using Feller's construction of all one-dimensional diffusions in Chapter 21. However, the construction of diffusions presented in Chapter 7 requires that $\sigma(\cdot)$ be Lipschitz. An alternative brief treatment of time change to

---

[3] Throughout, BCPT refers to Bhattacharya and Waymire (2016), A Basic Course in Probability Theory.

[4] See Ikeda and Watanabe (1989), pp. 85–86.

Brownian motion is given in McKean (1969) that does not require the martingale characterization of Brownian motion. Specifically, one has the following:

Theorem 9.6 *Let* $\mathbf{g}(s), s \geq 0$*, be an* $\mathbb{R}^k$ *valued nonanticipative process belonging to* $\mathcal{M}[0, \infty)$*, or* $\mathcal{M}[0, T]$ *for some* $T > 0$*. Assume that the quadratic variation* $q_t = \int_0^t |\mathbf{g}(s)|^2 ds, t \geq 0$*, of the (one-dimensional) process defined by the stochastic integral* $I(t) = \int_0^t \mathbf{g}(s) \cdot d\mathbf{B}_s, t \geq 0$*, is strictly increasing in* t*. Define the stopping times* $\tau_t = \inf\{s \geq 0 : q_s = t\}, t \geq 0$*. Then* $W_t = I(\tau_t), t \geq 0$*, is a standard one-dimensional Brownian motion with respect to the filtration* $\mathcal{G}_t = \mathcal{F}_{\tau_t}, t \geq 0$*.*

Proof For arbitrary $s < t$, and $\xi \in \mathbb{R}$, consider the process $C(t) = \exp\{i\xi I(t) + \frac{1}{2}\xi^2 q_t\}, t \geq 0$. By Itô's lemma, one has

$$\begin{aligned} dC(t) &= C(t)\{i\xi \mathbf{g}(t) \cdot d\mathbf{B}_t + \frac{1}{2}(i\xi|\mathbf{g}(t)|)^2 dt + (\frac{\xi^2}{2}|\mathbf{g}(t)|)^2 dt\} \\ &= C(t) i\xi \mathbf{g}(t) \cdot d\mathbf{B}_t, \end{aligned} \tag{9.8}$$

so that $C(t), t \geq 0$, is a $\mathcal{F}_t$-martingale. By the optional stopping theorem, it follows that $C(\tau_t), t \geq 0$, is a $\mathcal{G}_t$-martingale. Therefore, noting that $C(\tau_t) = \exp\{i\xi I(\tau_t) + \frac{1}{2}\xi^2 t\}$, one has for all $s < t$,

$$\mathbb{E} \exp\{i\xi I(\tau_t) + \frac{1}{2}\xi^2 t | \mathcal{G}_s\} = \exp\{i\xi I(\tau_s) + \frac{\xi^2}{2} s\}, \tag{9.9}$$

From this, one obtains that

$$\mathbb{E} \exp\{i\xi(I(\tau_t) - I(\tau_s)) | \mathcal{G}_s\} = \exp\{-\frac{\xi^2}{2}(t - s)\}, \tag{9.10}$$

proving that $I(\tau_t) - I(\tau_s)$ is independent of $\mathcal{G}_s$ and that it has the Gaussian distribution with mean zero and variance $t - s$. Moreover, $t \to W_t = I(\tau_t), t \geq 0$, is a composition of continuous functions and therefore continuous. ■

For an application of time change, one may obtain the following property of two-dimensional Brownian motion when viewed as a random path in the complex plane[5]

Corollary 9.7 *Let* $\mathbf{B} = (B^{(1)}, B^{(2)})$ *be a two-dimensional standard Brownian motion starting at the origin and let* $f \equiv u + iv : \mathbb{C} \to \mathbb{C}$ *be a nonconstant analytic function on the complex plane. Then* $(u(B), v(B))$ *is a time change of two-dimensional Brownian motion.*

Proof Observe that by the Cauchy-Riemann equations, namely,

[5] This result is generally attributed to Lévy (1948).

$$u_x = -v_y, u_y = v_x,$$

u and v are harmonic functions so that by Itô's lemma, $\{u(B_t) : t \geq 0\}$ and $\{v(B_t) : t \geq 0\}$ are stochastic integrals. Moreover, by the Cauchy-Riemann equations, $|\,\mathbf{grad}\, u|^2 = |\,\mathbf{grad}\, v|^2$ and $\mathbf{grad}\, u \cdot \mathbf{grad}\, v = 0$. Thus, the quadratic variation and therefore time change to Brownian motion are the same for both stochastic integrals. ■

Remark 9.5 Analytic functions f for which $f'(z) \neq 0$ for all $z \in \mathbb{C}$ are referred to as *conformal maps*. In this context, Corollary 9.7 is often cited as a form of a "conformal invariance"property of Brownian motion; see Lawler (2008) for a general perspective on the utility of conformal invariance in the context of scaling limits of discrete lattice systems.

Remark 9.6 One may observe that the martingale $v(B)$ induced by the imaginary part of $f(B)$ is represented by a stochastic integral that may be viewed as a linear transformation of the stochastic integral defining the martingale induced by the real part $u(B)$. Such "martingale transforms" play a probabilistic role related to that of Hilbert and Riesz transforms in harmonic analysis; see Exercise 3.

The following lemma is useful to note for computations in proofs to folllow.

Lemma 1 *Let* $\mathbf{B} = \{\mathbf{B}_t : t \geq 0\}$ *denote a k-dimensional standard Brownian motion on* $(\Omega, \mathcal{F}, P)$*, and let* $\boldsymbol{\gamma}(u) \equiv \boldsymbol{\gamma}(u, \cdot)$ *be a nonanticipative* $\mathbb{R}^k$*–valued functional on* $[0, T]$ *such that* $\mathbb{E}\exp\{\theta \int_0^T |\boldsymbol{\gamma}(u)|^2 du\} < \infty$ *for some* $\theta > 1$*. Define*

$$M_t := \exp\{\int_0^t \boldsymbol{\gamma}(u) \cdot d\mathbf{B}_u - \frac{1}{2}\int_0^t |\boldsymbol{\gamma}(u)|^2 du\}, 0 \leq t \leq T.$$

Then under P*, one has*

$$dM_t = M_t \boldsymbol{\gamma}(t) \cdot d\mathbf{B}_t, \tag{9.11}$$

i.e., $M_t = 1 + \int_0^t M_s \boldsymbol{\gamma}(s) \cdot d\mathbf{B}_s, t \geq 0$*, is a martingale.*

Proof By Itô's lemma, under P, one has

$$dM_t = M_t[\boldsymbol{\gamma}(t) \cdot d\mathbf{B}_t - \frac{1}{2}|\boldsymbol{\gamma}(t)|^2 dt] + \frac{1}{2}M_t|\boldsymbol{\gamma}(t)|^2 dt = M_t \boldsymbol{\gamma}(t) \cdot d\mathbf{B}_t. \tag{9.12}$$

Thus, the assertion follows. ■

The next corollary paves the way to prove the main result of this chapter.

Corollary 9.8 *Let* $\mathbf{B} = \{\mathbf{B}_t : t \geq 0\}$ *denote a k-dimensional standard Brownian motion on* $(\Omega, \mathcal{F}, P)$*, and let* $\boldsymbol{\gamma}(u) \equiv \boldsymbol{\gamma}(u, \cdot)$ *be a nonanticipative* $\mathbb{R}^k$*-valued functional on* $[0, T]$ *such that* $\mathbb{E}\exp\{\theta \int_0^T |\boldsymbol{\gamma}(u)|^2 du\} < \infty$ *for some* $\theta > 1$*, and define* $M_t := \exp\{\int_0^t \boldsymbol{\gamma}(u) \cdot d\mathbf{B}_u - \frac{1}{2}\int_0^t |\boldsymbol{\gamma}(u)|^2 du\}$*,* $0 \leq t \leq T$*.*

a. *Then a probability measure on $\mathcal{F}_T$ is defined by*

$$Q(A) = \mathbb{E}\mathbf{1}_A M_T \equiv \int_A M_T dP \qquad \forall\ A \in \mathcal{F}_T, \tag{9.13}$$

and one has

$$Q(A) = \mathbb{E}\mathbf{1}_A M_t \qquad \forall\ A \in \mathcal{F}_t,\ 0 \le t \le T, \tag{9.14}$$

i.e., denoting the respective restrictions of Q, P to $\mathcal{F}_t$ by $Q_{[0,t]}$, $P_{[0,t]}$ one has $\frac{dQ_{[0,t]}}{dP_{[0,t]}} = M_t$ on $\mathcal{F}_t$.

b. *Under Q,*

$$\tilde{\mathbf{B}}_t := \mathbf{B}_t - \int_0^t \boldsymbol{\gamma}(u)du, \qquad 0 \le t \le T,$$

is a standard k-dimensional Brownian motion.

Proof

a. That Q is a probability measure on $\mathcal{F}_T$ follows from the fact that $M_t, 0 \le t \le T$, is a martingale so that $\mathbb{E}M_T = \mathbb{E}M_0 = 1$. Also, if $A \in \mathcal{F}_t$ $(0 \le t \le T)$, then $\mathbb{E}\mathbf{1}_A M_T = \mathbb{E}[\mathbf{1}_A M_t]$, by the martingale property.

b. By Theorem 9.4, we need to prove that, under Q, (1) $\widetilde{B}_t^{(i)}$ $(1 \le i \le k)$, $0 \le t \le T$, are mean-zero martingales and (2) $\widetilde{B}_t^{(i)}\widetilde{B}_t^{(j)} - \delta_{ij}t$ $(0 \le t \le T)$ are martingales. For $A \in \mathcal{F}_s$, $s < t \le T$, writing $\mathbb{E}_Q$ and $\mathbb{E}_P$ $(= \mathbb{E})$ for expectations under Q and P, respectively, on $\mathcal{F}_T$, we have

$$\mathbb{E}_Q(\mathbf{1}_A\widetilde{B}_t^{(i)}) = \mathbb{E}_P(\mathbf{1}_A\widetilde{B}_t^{(i)}M_t) = \mathbb{E}_P[\mathbf{1}_A\mathbb{E}_P(\widetilde{B}_t^{(i)}M_t \mid \mathcal{F}_s)]. \tag{9.15}$$

Using Lemma 1, one has under P that $M_t = 1 + \int_0^t M_s\boldsymbol{\gamma}(s)\cdot d\mathbf{B}_s$, and

$$\begin{aligned}
&d(\widetilde{B}_t^{(i)}M_t)\\
&= d[(B_t^{(i)} - \int_0^t \gamma^{(i)}(u)du)M_t]\\
&= d(B_t^{(i)}M_t) - \gamma^{(i)}(t)M_t dt - (\int_0^t \gamma^{(i)}(u)du)dM_t\\
&= M_t dB_t^{(i)} + B_t^{(i)}dM_t + M_t\gamma^{(i)}(t)dt - \gamma^{(i)}(t)M_t dt - (\int_0^t \gamma^{(i)}(u)du)dM_t\\
&= (B_t^{(i)} - \int_0^t \gamma^{(i)}(u)du)dM_t + M_t dB_t^{(i)}.
\end{aligned} \tag{9.16}$$

Hence, under P, $\widetilde{B}_t^{(i)} M_t$ is a stochastic integral and, therefore, a (mean-zero) martingale. Using this in (9.15), one gets $\forall\, A \in \mathcal{F}_s$,

$$\mathbb{E}_Q(\mathbf{1}_A \widetilde{B}_t^{(i)}) = \mathbb{E}_P(\mathbf{1}_A \widetilde{B}_s^{(i)} M_s) = \mathbb{E}_Q(\mathbf{1}_A \widetilde{B}_s^{(i)}), \tag{9.17}$$

so that $\widetilde{B}_t^{(i)}$, $0 \le t \le T$, is a mean-zero martingale ($1 \le i \le k$). Next note that under P, using Itô's lemma,

$$\begin{aligned} & d[(\widetilde{B}_t^{(i)} \widetilde{B}_t^{(j)} - \delta_{ij} t) M_t] \\ &= (\widetilde{B}_t^{(i)} \widetilde{B}_t^{(j)} - \delta_{ij} t) dM_t + (\widetilde{B}_t^{(i)} d\widetilde{B}_t^{(j)} + \widetilde{B}_t^{(j)} d\widetilde{B}_t^{(i)} + \delta_{ij} - \delta_{ij}) M_t \\ &= (\widetilde{B}_t^{(i)} \widetilde{B}_t^{(j)} - \delta_{ij} t) dM_t + M_t (\widetilde{B}_t^{(i)} d\widetilde{B}_t^{(j)} + \widetilde{B}_t^{(j)} d\widetilde{B}_t^{(i)}). \end{aligned}$$

Thus, under P, $(\widetilde{B}_t^{(i)} \widetilde{B}_t^{(j)} - \delta_{ij} t) M_t$ is a stochastic integral and, therefore, a martingale. Therefore, as in the proof of (9.17), using (9.16), $\widetilde{B}_t^{(i)} \widetilde{B}_t^{(j)} - \delta_{ij} t$, $0 \le t \le T$, is a martingale under Q. ■

We are now ready to state and prove the CMG Theorem. Consider the diffusion

$$\mathbf{X}_t^{\mathbf{x},0} = \mathbf{x} + \int_0^t \boldsymbol{\mu}(u, \mathbf{X}_u^{\mathbf{x},0}) du + \int_0^t \boldsymbol{\sigma}(u, \mathbf{X}_u^{\mathbf{x},0}) d\mathbf{B}_u, \tag{9.18}$$

where $\boldsymbol{\mu}(u, \mathbf{y})$, $\boldsymbol{\sigma}(u, \mathbf{y})$ are measurable (on $[0, T] \times \mathbb{R}^k$) and spatially Lipschitzian, uniformly over $u \in [0, T]$. Let $\boldsymbol{\gamma}(u, \mathbf{y})$ be measurable on $[0, T] \times \mathbb{R}^k$ into $\mathbb{R}^k$, spatially Lipschitzian (uniformly on $[0, T]$) and such that $\boldsymbol{\sigma}(u, \mathbf{y})\boldsymbol{\gamma}(u, \mathbf{y})$ is also spatially Lipschitzian uniformly on $[0, T]$. Define

$$M_t := \exp\{\int_0^t \boldsymbol{\gamma}(u, \mathbf{X}_u^{\mathbf{x},0}) \cdot d\mathbf{B}_u - \frac{1}{2} \int_0^t |\boldsymbol{\gamma}(u, \mathbf{X}_u^{\mathbf{x},0})|^2 du\}, \qquad 0 \le t \le T. \tag{9.19}$$

Let Q be the probability measure on $\mathcal{F}_T$ defined by

$$Q(A) = \int_A M_T dP \qquad (A \in \mathcal{F}_T). \tag{9.20}$$

Theorem 9.9 (Cameron–Martin–Girsanov Theorem (CMG)) *Assume that for some $\theta > 1$, $\mathbb{E} \exp\{\theta \int_0^T |\boldsymbol{\gamma}(u, \mathbf{X}_u^{\mathbf{x},0})|^2 du\} < \infty$. Under the above notation and hypotheses (9.18)–(9.20), the following are true.*

a. *Under Q,*

$$\widetilde{\mathbf{B}}_t := \mathbf{B}_t - \int_0^t \boldsymbol{\gamma}(u, \mathbf{X}_u^{\mathbf{x},0}) du, \qquad 0 \le t \le T, \tag{9.21}$$

is a standard $\{\mathcal{F}_t : t \geq 0\}$–Brownian motion on $\mathbb{R}^k$ (on the interval of time $[0, T]$).

b. *Under Q, $\mathbf{X}_u^{\mathbf{x},0}$ $(0 \leq u \leq T)$ is a diffusion with drift $\boldsymbol{\mu}(u, \mathbf{y}) + \boldsymbol{\sigma}(u, \mathbf{y})\boldsymbol{\gamma}(u, \mathbf{y})$ and diffusion coefficient $\boldsymbol{\sigma}(u, \mathbf{y})$.*

c. *The distributions $P_T^{\mathbf{x},0}$ and $Q_T^{\mathbf{x},0}$, say of $X^{\mathbf{x},0}$ under P and Q, respectively, are mutually absolutely continuous and related by (9.23).*

Proof

(a) is a special case of Corollary 9.8. Part (b) follows on rewriting (9.18) as

$$\mathbf{X}_t^{\mathbf{x},0} = \mathbf{x} + \int_0^t \{\boldsymbol{\mu}(u, \mathbf{X}_u^{\mathbf{x},0}) + \boldsymbol{\sigma}(u, \mathbf{X}_u^{\mathbf{x},0})\boldsymbol{\gamma}(u, \mathbf{X}_u^{\mathbf{x},0})\}du + \int_0^t \boldsymbol{\sigma}(u, \mathbf{X}_u^{\mathbf{x},0})d\widetilde{\mathbf{B}}_u, \quad (0 \leq t \leq T), \quad (9.22)$$

noting that, under Q, $\widetilde{\mathbf{B}}_u$ $0 \leq u \leq T$, is a standard k-dimensional $\{\mathcal{F}_t : t \geq 0\}$-Brownian motion. To prove part (c), note that for every Borel subset D of $C([0, T] : \mathbb{R}^k)$, the set $A := [\mathbf{X}_{[0,T]}^{\mathbf{x},0} \in D] \in \mathcal{F}_T$, where $\mathbf{X}_{[0,T]}^{\mathbf{x},0}(\omega)$ is the function $t \to \mathbf{X}_t^{\mathbf{x},0}(\omega)$, $0 \leq t \leq T$, $\forall\, \omega \in \Omega$. Hence,

$$Q_T^{\mathbf{x},0}(D) = Q(A) = \int_A M_T dP, \qquad P_T^{\mathbf{x},0}(D) = P(A), \tag{9.23}$$

so that $Q_T^{\mathbf{x},0}(D) = 0$ if and only if $P_T^{\mathbf{x},0}(D) = 0$.

■

Corollary 9.10 *Let the vector fields (of $\mathbb{R}^k$ into $\mathbb{R}^k$) $\boldsymbol{\mu}(t, \mathbf{y})$ and $\boldsymbol{\beta}(t, \mathbf{y})$ be spatially Lipschitz uniformly for $t \in [0, T]$. Also assume $\boldsymbol{\sigma}(t, \mathbf{y})$ is spatially Lipschitz uniformly for $t \in [0, T]$ and that it is nonsingular such that*

$$\boldsymbol{\gamma}(t, \mathbf{y}) := \boldsymbol{\sigma}^{-1}(t, \mathbf{y})(\boldsymbol{\beta}(t, \mathbf{y}) - \boldsymbol{\mu}(t, \mathbf{y})), \qquad 0 \leq t \leq T, \tag{9.24}$$

is bounded. Then the distributions on $C([0, T) : \mathbb{R}^k)$ of the two diffusions with drifts $\boldsymbol{\mu}(t, \mathbf{y})$ and $\boldsymbol{\beta}(t, \mathbf{y})$, respectively, and having the same diffusion coefficient $\boldsymbol{\sigma}(t, \mathbf{y})$, are mutually absolutely continuous.

Corollary 9.10 is an immediate consequence of part (c) of Theorem 9.9.

Remark 9.7 Note that the vector field $\boldsymbol{\gamma}(u)$ in Corollary 9.8 is only required to be nonanticipative, and need not be continuous or Lipschitzian, as long as the integrability condition is satisfied (e.g., if $\boldsymbol{\gamma}$ is bounded on $[0, T]$). Similarly, part (a) of Theorem 9.9 holds if $\boldsymbol{\gamma}(u, \mathbf{X}_u^{\mathbf{x},0})$ is nonanticipative and bounded on $[0, T] \times \Omega$. Let now $\boldsymbol{\mu}(u, \mathbf{y}) \equiv 0$, $\boldsymbol{\sigma}(u, \mathbf{y})$ spatially Lipschitz (uniformly on $[0, T]$), and $\boldsymbol{\gamma}(u, \mathbf{y})$ bounded measurable on $[0, T] \times \mathbb{R}^k$ (into $\mathbb{R}^k$). Then the representation (9.22) holds under Q. However, its Markov property is not immediately clear, since the conditions for the pathwise uniqueness of the solution to an *equation* such

as (9.22) are not met. The Markov property was proved by Stroock and Varadhan (1969), who showed that parts (b) and (c) still hold; indeed, one requires only that $\boldsymbol{\sigma}(u, \mathbf{y}), \boldsymbol{\sigma}^{-1}(u, \mathbf{y})$ are continuous and bounded. The same argument leads to the conclusion of Corollary 9.10 under the same assumption on $\boldsymbol{\sigma}$, and letting $\boldsymbol{\mu} = 0$ and $\boldsymbol{\beta}(u, \mathbf{y})$ bounded and measurable. In particular, one may define a diffusion on $\mathbb{R}^k$ with drift $\boldsymbol{\mu}(u, \mathbf{y})$ bounded and measurable and diffusion $\boldsymbol{\sigma}(u, \mathbf{y})$ continuous and *uniformly elliptic* (i.e., $\boldsymbol{\sigma}, \boldsymbol{\sigma}^{-1}$ both bounded). This result is quite deep (see Stroock and Varadhan (1979)).

We now discuss the second approach to the CMG Theorem following McKean (1969). Assume $\boldsymbol{\mu}(\cdot), \boldsymbol{\sigma}(\cdot)$ locally Lipschitzian, $\boldsymbol{\sigma}(\mathbf{x})$ nonsingular for all $\mathbf{x}$. Consider the diffusion $\mathbf{X}^{\mathbf{x}_0}$ on $\mathbb{R}^k$ governed by

$$\mathbf{X}_t^{\mathbf{x}_0} = \mathbf{x}_0 + \int_0^t \boldsymbol{\sigma}(\mathbf{X}_s^{\mathbf{x}_0}) d\mathbf{B}_s, \tag{9.25}$$

where $\{\mathbf{B}_t : t \geq 0\}$ is a standard k-dimensional Brownian motion with respect to a filtration $\{\mathcal{F}_t : t \geq 0\}$ on a probability space. Assume that this diffusion is nonexplosive. This is automatic for $k = 1$ by the tests of Feller in Chapter 12 but requires additional conditions on $\boldsymbol{\sigma}\boldsymbol{\sigma}'$ in higher dimensions to be provided in Chapter 12. Let $\mathbf{Y}^{\mathbf{x}_0}$ be the diffusion governed by the Itô equation

$$\mathbf{Y}_t^{\mathbf{x}_0} = \mathbf{x}_0 + \int_0^t \boldsymbol{\mu}(\mathbf{Y}_s^{\mathbf{x}_0}) ds + \int_0^t \boldsymbol{\sigma}(\mathbf{Y}_s^{\mathbf{x}_0}) d\mathbf{B}_s. \tag{9.26}$$

Let ζ denote the explosion time of $\mathbf{Y}^{\mathbf{x}_0}$. Write

$$M_{t,\mathbf{x}_0}^0 := \exp\{\int_0^t \boldsymbol{\sigma}^{-1}(\mathbf{X}_s^{\mathbf{x}_0})\boldsymbol{\mu}(\mathbf{X}_s^{\mathbf{x}_0}) \cdot d\mathbf{B}_s - \frac{1}{2}\int_0^t |\boldsymbol{\sigma}^{-1}(\mathbf{X}_s^{\mathbf{x}_0})\boldsymbol{\mu}(\mathbf{X}_s^{\mathbf{x}_0})|^2 ds\}. \tag{9.27}$$

Below "distribution," without qualification, refers to distribution under P, and $\mathbb{E}$ denotes expectation under P.

Theorem 9.11 (Cameron–Martin–Girsanov Theorem (CMG)) *Under the above hypothesis, one has, for each $t \geq 0$,*

$$P(\mathbf{Y}_{[0,t]}^{\mathbf{x}_0} \in C, \zeta > t) = \mathbb{E}(\mathbf{1}_{[\mathbf{X}_{[0,t]}^{\mathbf{x}_0} \in C]} M_{t,\mathbf{x}_0}^0) \tag{9.28}$$

for every Borel subset C of $C([0,t] : \mathbb{R}^k)$. Here, the subscript $[0,t]$ indicates restriction of the processes to the time interval $[0,t]$.

Proof Define, for each positive integer $n > |\mathbf{x}_0|$,

$$\tau_n := \inf\{t \geq 0 : |\mathbf{X}_t^{\mathbf{x}_0}| = n\}, \quad \tilde{\tau}_n := \inf\{t \geq 0 : |\mathbf{Y}_t^{\mathbf{x}_0}| = n\}. \tag{9.29}$$

It is enough to prove that, for each n, and each $t > 0$,

$$P(\mathbf{Y}^{\mathbf{x}_0}_{[0,t]} \in C,\ \tilde{\tau}_n > t) = \mathbb{E}(\mathbf{1}_{[\mathbf{X}^{\mathbf{x}_0}_{[0,t]} \in C] \cap [\tau_n > t]} M_t). \tag{9.30}$$

To prove (9.30), one may arbitrarily modify $\boldsymbol{\mu}(\cdot)$, $\boldsymbol{\sigma}(\cdot)$ such that $\boldsymbol{\mu}(\mathbf{x})$ vanishes for sufficiently large values of $|\mathbf{x}|$, and $\boldsymbol{\sigma}(\mathbf{x}) = I$ for sufficiently large values of $|\mathbf{x}|$, and $\boldsymbol{\mu}(\cdot)$, $\boldsymbol{\sigma}(\cdot)$ are bounded and globally Lipschitzian. With this modification, define for $0 \le s \le t$,

$$M^s_{t,\mathbf{x}_0} := \exp\{\int_s^t \boldsymbol{\sigma}^{-1}(\mathbf{X}^{\mathbf{x}_0}_u)\boldsymbol{\mu}(\mathbf{X}^{\mathbf{x}_0}_u) d\mathbf{B}_u - \frac{1}{2}\int_s^t |\boldsymbol{\sigma}^{-1}(\mathbf{X}^{\mathbf{x}_0}_u)\boldsymbol{\mu}(\mathbf{X}^{\mathbf{x}_0}_u)|^2 du\}. \tag{9.31}$$

Then $M^0_{s,\mathbf{x}_0}$, $s \ge 0$, is a $\{\mathcal{F}_s\}_{s \ge 0}$-martingale, by Proposition 9.1, and $M^0_{s,\mathbf{x}_0} = 1 + \int_0^t M^0_{u,\mathbf{x}_0}\boldsymbol{\sigma}^{-1}(\mathbf{X}^{\mathbf{x}_0}_u)\boldsymbol{\mu}(\mathbf{X}^{\mathbf{x}_0}_u) \cdot d\mathbf{B}_u$. As before, define the probability measure $Q_{\mathbf{x}_0}$ on $\mathcal{F}_t$ by

$$Q_{\mathbf{x}_0}(A) = \mathbb{E}(\mathbf{1}_A M^0_{t,\mathbf{x}_0}) \qquad \forall\ A \in \mathcal{F}_t. \tag{9.32}$$

We will first show that, under $\mathbf{Q}_{\mathbf{x}_0}$, $X^{\mathbf{x}_0}_s$, $0 \le s \le t$, is a Markov process with respect to $\{\mathcal{F}_s\}_{0 \le s \le t}$. Let G be a real bounded $\mathcal{F}_s$-measurable (test) function and g a real bounded Borel measurable function on $\mathbb{R}^k$. Then

$$\begin{aligned}
\mathbb{E}_{Q_{\mathbf{x}_0}} G g(\mathbf{X}^{\mathbf{x}_0}_{s+u}) &= \mathbb{E} G g(\mathbf{X}^{\mathbf{x}_0}_{s+u}) M^0_{s+u,\mathbf{x}_0} \\
&= \mathbb{E}[G M^0_{s,\mathbf{x}_0} g(\mathbf{X}^{\mathbf{x}_0}_{s+u}) M^s_{s+u,\mathbf{x}_0}] \\
&= \mathbb{E}[G M^0_{s,\mathbf{x}_0} \mathbb{E}(g(\mathbf{X}^{\mathbf{x}_0}_{s+u}) M^s_{s+u,\mathbf{x}_0} \mid \mathcal{F}_s) \\
&= \mathbb{E}[G M^0_{s,\mathbf{x}_0} \{\mathbb{E}(g(\mathbf{X}^{\mathbf{y}}_u) M^0_{u,\mathbf{y}}\}_{\mathbf{y}=\mathbf{X}^{\mathbf{x}_0}_s}],
\end{aligned} \tag{9.33}$$

since (i) $\{\mathbf{B}_{s+u} - \mathbf{B}_s, 0 \le u \le t-s\}$ is independent of $\mathcal{F}_s$ and (ii) $(\mathbf{X}^{\mathbf{x}_0}_{s+u}, M^s_{s+u,\mathbf{x}_0})$ is the same function of $(\mathbf{X}^{\mathbf{x}_0}_s, \{\mathbf{B}_{s+u'} - \mathbf{B}_s, 0 \le u' \le u\})$ as $(\mathbf{X}^{\mathbf{y}}_u, M^0_{u,\mathbf{y}})$ is of $(\mathbf{y}, \{\mathbf{B}_{u'}, 0 \le u' \le u\})$. The last expression of (9.33) equals

$$\mathbb{E}[G M^0_{s,\mathbf{x}_0}(\mathbb{E}_{Q_\mathbf{y}} g(\mathbf{X}^{\mathbf{y}}_u))_{\mathbf{y}=\mathbf{X}^{\mathbf{x}_0}_s}] = \mathbb{E}_{Q_{\mathbf{x}_0}}[G(\mathbb{E}_{Q_\mathbf{y}} g(\mathbf{X}^{\mathbf{y}}_u))_{\mathbf{y}=\mathbf{X}^{\mathbf{x}_0}_s}].$$

That is, under $Q_{\mathbf{x}_0}$, the conditional distribution of $\mathbf{X}^{\mathbf{x}_0}_{s+u}$, given $\mathcal{F}_s$ is the $Q_\mathbf{y}$-distribution of $\mathbf{X}^{\mathbf{y}}_u$ on $[X^{\mathbf{x}_0}_s = \mathbf{y}]$. This proves the Markov property of $\mathbf{X}^{\mathbf{x}}_s$, $0 \le s \le t$, under $Q_{\mathbf{x}_0}$. It remains to prove that the $Q_{\mathbf{x}_0}$-distribution of $\mathbf{X}^{\mathbf{x}_0}_s$ $0 \le s \le t$, is the same as the distribution of $\mathbf{Y}^{\mathbf{x}_0}_s$, $0 \le s \le t$. For this, let g be a bounded function on $\mathbb{R}^k$ with bounded and continuous first-order derivatives. Then, by Itô's lemma, and recalling Lemma 1,

$$
\begin{aligned}
\mathbb{E}_{Q_{\mathbf{x}_0}} g(\mathbf{X}_s^{\mathbf{x}_0}) - g(\mathbf{x}_0) &= \mathbb{E}[g(\mathbf{X}_s^{\mathbf{x}_0}) M_{s,\mathbf{x}_0}^0] - g(\mathbf{x}_0) \\
&= \mathbb{E}[(\int_0^t M_{u,\mathbf{x}_0}^0 \boldsymbol{\sigma}(X_u^{\mathbf{x}_0}) \nabla g(\mathbf{X}_u^{\mathbf{x}_0}) \cdot d\mathbf{B}_u) \\
&\quad + \frac{1}{2} \int_0^t \sum (\partial_{rr'} g)(\mathbf{X}_u^{\mathbf{x}_0}) \sigma_{rr'}(\mathbf{X}_u^{\mathbf{x}_0}) M_{u,\mathbf{x}_0}^0 du \\
&\quad + \int_0^t M_{u,\mathbf{x}_0}^0 g(\mathbf{X}_u^{\mathbf{x}_0}) \boldsymbol{\sigma}^{-1}(\mathbf{X}_u^{\mathbf{x}_0}) \boldsymbol{\mu}(\mathbf{X}_u^{\mathbf{x}_0}) \cdot d\mathbf{B}_u \\
&\quad + \int_0^t (\boldsymbol{\sigma}(X_u^{\mathbf{x}_0}) \nabla g(\mathbf{X}_u^{\mathbf{x}_0}) \cdot \boldsymbol{\sigma}^{-1}(\mathbf{X}_u^{\mathbf{x}_0}) \boldsymbol{\mu}(\mathbf{X}_u^{\mathbf{x}_0})) M_{u,\mathbf{x}_0}^0 du] \\
&= \mathbb{E}[\int_0^s (Ag)(\mathbf{X}_u^{\mathbf{x}_0}) M_{u,\mathbf{x}_0}^0 du] = \int_0^s [\mathbb{E}_{Q_{\mathbf{x}_0}} Ag(\mathbf{X}_u^{\mathbf{x}_0})] du
\end{aligned}
$$

where A is the generator of the diffusion $\mathbf{Y}_s^{\mathbf{x}_0}$, $0 \le s \le t$. This shows that the $Q_{\mathbf{x}_0}$-distribution of $X_s^{\mathbf{x}_0}$, $0 \le s \le t$, is the same as the distribution (under P) of $\mathbf{Y}_s^{\mathbf{x}_0}$, $0 \le s \le t$, since the two Markov processes have the same generator. Since $t \ge 0$ is arbitrary, the theorem is proved. ■

Corollary 9.12 *Under the hypothesis of Theorem 9.11,*

a. *The distribution of the explosion time ζ of $\mathbf{Y}^{\mathbf{x}_0}$ is given by*

$$P(\zeta > t) = \mathbb{E} M_{t,\mathbf{x}_0}^0, \tag{9.34}$$

so that

b. *$\mathbf{Y}_t^{\mathbf{x}_0}$ is nonexplosive if and only if $\mathbb{E} M_{t,\mathbf{x}_0}^0 = 1$ for all t (or, equivalently, $\{M_{t,\mathbf{x}_0}^0, t \ge 0\}$ is a martingale) for every $\mathbf{x}_0 \in \mathbb{R}^k$, and in this case*

c. *The distribution of $\mathbf{Y}_{[0,T]}^{\mathbf{x}_0}$ is absolutely continuous with respect to the distribution of $\mathbf{X}_{[0,T]}^{\mathbf{x}_0}$ for all $T > 0$, and all $\mathbf{x}_0$.*

Remark 9.8 Since necessary and sufficient conditions are known for nonexplosion in one dimension, Corollary 9.12 provides a characterization of the process $M_{t,\mathbf{x}_0}^0$, $t \ge 0$ ($\mathbf{x}_0 \in \mathbb{R}^k$) being a martingale. The use of Novikov's condition here follows as a consequence. For $k > 1$, one may use Khaśminskii's sufficient condition for nonexplosion to arrive at a criterion for the martingale property of $M_{t,\mathbf{x}_0}^0$, $t \ge 0$ more general than what an application of Novikov's condition yields (see Chapter 12).

In Theorem 9.11, the diffusion $\mathbf{X}_t$ is taken to be driftless. However, one can use a general nonexplosive diffusion

$$\mathbf{X}_t^{\mathbf{x}_0} = \mathbf{x}_0 + \int_0^t \boldsymbol{\beta}(\mathbf{X}_s^{\mathbf{x}_0}) ds + \int_0^t \boldsymbol{\sigma}(\mathbf{X}_s^{\mathbf{x}_0}) d\mathbf{B}_s \tag{9.35}$$

and arrive at the following extension, more or less using the same arguments (Exercise 8).

Corollary 9.13 *Assume $\boldsymbol{\beta}(\cdot)$ and $\boldsymbol{\sigma}(\cdot)$ are locally Lipschitzian, and $\boldsymbol{\sigma}(\mathbf{x})$ is nonsingular for all $\mathbf{x}$, and that $\mathbf{X}^{\mathbf{x}}$ ($\mathbf{x} \in \mathbb{R}^k$) is nonexplosive. Then if $\boldsymbol{\mu}(\cdot)$ is locally Lipschitzian, one has for every $t > 0$, $\mathbf{x}_0 \in \mathbb{R}^k$,*

$$P(\mathbf{Y}^{\mathbf{x}_0}_{[0,t]} \in C, \zeta > t) = \mathbb{E}(\mathbf{1}_{[\mathbf{X}^{\mathbf{x}_0}_{[0,t]} \in C]} M^0_{t,\mathbf{x}_0}) \tag{9.36}$$

where C is a Borel subset of $C([0,t]:\mathbb{R}^k)$ and

$$M^0_{t,\mathbf{x}_0} := \exp\{\int_0^t \boldsymbol{\sigma}^{-1}(\mathbf{X}^{\mathbf{x}_0}_s)[\boldsymbol{\mu}(\mathbf{X}^{\mathbf{x}_0}_s) - \boldsymbol{\beta}(\mathbf{X}^{\mathbf{x}_0}_s)] \cdot d\mathbf{B}_s$$

$$-\frac{1}{2}\int_0^t |\boldsymbol{\sigma}^{-1}(\mathbf{X}^{\mathbf{x}_0}_s)[\boldsymbol{\mu}(\mathbf{X}^{\mathbf{x}_0}_s) - \boldsymbol{\beta}(\mathbf{X}^{\mathbf{x}_0}_s)]|^2 ds\}. \tag{9.37}$$

Example 2 (First Passage Time Under Drift) Recall that the hitting time τ_y of a point y by a one-dimensional standard Brownian motion $\{B_t : t \geq 0\}$ starting at zero has distribution that may be readily computed from the reflection principle (strong Markov property),

$$P(\max_{0 \leq s \leq t} B_s > y) = 2P(B_t > y), \quad y > 0,$$

to obtain the probability density

$$f_y(t) := \frac{|y|}{\sqrt{2\pi t^3}} e^{-\frac{y^2}{2t}}.$$

This calculation extends to Brownian motion with drift μ by an application of CMG and the optional stopping theorem with respect to the pair of stopping times $\tau_y \wedge t \leq t$ as follows. Applying CMG, conditioning on $\mathcal{F}_{\tau_y \wedge t}$ and then using the optional stopping rule, one has

$$\begin{aligned} P(\inf\{s : B_s + \mu s = y\} \leq t) &= \mathbb{E}(\mathbf{1}_{[\tau_y \leq t]} M_t) = \mathbb{E}(\mathbf{1}_{[\tau_y \leq t]} \mathbb{E}(M_t | \mathcal{F}_{\tau_y \wedge t})) \\ &= \mathbb{E}(\mathbf{1}_{[\tau_y \leq t]} M_{\tau_y \wedge t}). \end{aligned} \tag{9.38}$$

Now, on $[\tau_y \leq t]$ one has $M_{\tau_y \wedge t} = M_{\tau_y}$. Thus,

$$\begin{aligned} P(\inf\{s : B_s + \mu s = y\} \leq t) = \mathbb{E}(\mathbf{1}_{[\tau_y \leq t]} M_{\tau_y}) &= \int_0^t e^{\mu y - \frac{1}{2}\mu^2 s} \frac{|y|}{\sqrt{2\pi s^3}} e^{-\frac{y^2}{2s}} ds \\ &= \int_0^t \frac{|y|}{\sqrt{2\pi s^3}} e^{-\frac{(y-\mu s)^2}{2s}} ds. \end{aligned} \tag{9.39}$$

Example 3 (Boundary Value Distribution of Brownian Motion on a Half-Space) Let $\mathbf{B} = \{\mathbf{B}_t : t \geq 0\}$ be a k-dimensional standard Brownian motion starting at $\mathbf{x} \in H_k := \{\mathbf{x} : x_k > 0\}$, $(k \geq 2)$. H_k is referred to as the upper half-space in $\mathbb{R}^k$. Let $\tau = \inf\{t : \mathbf{B}_t \in \partial H_k\}$ where $\partial H_k = \{\mathbf{x} \in \mathbb{R}^k : x_k = 0\}$. Since the coordinates of $\mathbf{B}$ are independent, standard Brownian motions, τ is simply the first passage time of standard Brownian motion $B^{(k)}$ to 0 starting from $x_k > 0$. Let us compute the distribution of $\mathbf{B}_\tau = (B_\tau^{(1)}, \ldots, B_\tau^{(k-1)}, 0)$. For a Borel measurable set $A = A_{k-1} \times \{0\} \subseteq \partial H_k$, note that that $P(\mathbf{B}_\tau \in A) = \mathbb{E}\psi((B^{(1)}, \ldots, B^{(k-1)}), \tau)$ where $\psi(\mathbf{g}, t) = \mathbf{1}_{A_{k-1}}(\mathbf{g}(t))$ for a continuous function $\mathbf{g} : C[0, \infty) \to \mathbb{R}^{k-1}$, and real number $t \geq 0$. Since τ is a function of $B^{(k)}$, it is independent of $(B^{(1)}, \ldots, B^{(k-1)})$. Thus, one may apply the substitution property of conditional expectation to write

$$P(\mathbf{B}_\tau \in A) = \mathbb{E}[\mathbb{E}\{\psi((B^{(1)}, \ldots, B^{(k-1)}, 0), \tau)|\sigma(\tau)\}]$$

where, writing $\mathbf{x}^{k-1} = (x_1, \ldots, x_{k-1})$,

$$\begin{aligned}\mathbb{E}\psi((B^{(1)}, \ldots, B^{(k-1)}), \tau)|\tau] &= P[(B_t^{(1)}, \ldots, B_t^{(k-1)}) \in A_{k-1}]|_{t=\tau} \\ &= \int_{A_{k-1}} (\frac{1}{\sqrt{2\pi\tau}})^{k-1} e^{-\frac{1}{2\pi\tau}|\mathbf{y}-\mathbf{x}^{k-1}|^2} d\mathbf{y}.\end{aligned}$$

Thus, applying Fubini-Tonneli and making a change of variable $s = (|\mathbf{y} - \mathbf{x}^{k-1}|^2 + x_k^2)/2t$ one has

$$\begin{aligned}&P(\mathbf{B}_\tau \in A) \\ &= \int_{A_{k-1}} \int_0^\infty (\frac{1}{\sqrt{2\pi}})^k x_k \frac{(|\mathbf{y} - \mathbf{x}^{k-1}|^2 + x_k^2)}{2t^2} (\frac{2t}{|\mathbf{y} - \mathbf{x}^{k-1}|^2 + x_k^2})^{\frac{k+2}{2}} e^{-t} dt d\mathbf{y} \\ &= \int_{A_{k-1}} \Gamma(\frac{k}{2}) \pi^{-\frac{k}{2}} \frac{x_k}{(|\mathbf{y} - \mathbf{x}^{k-1}|^2 + x_k^2)^{\frac{k}{2}}} d\mathbf{y}. \qquad (9.40)\end{aligned}$$

The probability density function of $(B_\tau^{(1)}, \ldots, B_\tau^{(k-1)})$ can be read off from here. Note that in the case $k = 2$, this is a Cauchy density (also see Exercise 3).

Exercises

1. (*Elementary Change of Measure*) Suppose that Z is a standard normal random variable defined on $(\Omega, \mathcal{F}, P)$, and let $X = \sigma Z, Y = \sigma Z + \mu$. Then X, Y are random variables on $(\Omega, \mathcal{F}, P)$ with densities $f_a(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, x \in \mathbb{R}, a = 0, a = \mu$, respectively.

(a) Show that the probability distributions P_X and P_Y on $\mathbb{R}$ are mutually absolutely continuous with $\frac{dP_Y}{dP_X}(x) = \exp(\frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2})$.[*Hint:* Write $P_Y(B) = \int_B f_Y(y)dy = \int_B \frac{f_Y(x)}{f_X(x)} f_X(x)dx$, $B \in \mathcal{B}$.]
(b) Show $P(Y \in B) = \mathbb{E}\mathbf{1}(Y \in B) = \mathbb{E}(\mathbf{1}(X \in B)M)$, where $M = \exp(\frac{\mu}{\sigma^2}X - \frac{\mu^2}{2\sigma^2})$.
(c) Define a probability Q on $(\Omega, \mathcal{F}, P)$ by $Q(A) = \int_A M(\omega)P(d\omega)$, $A \in \mathcal{F}$. Show that under Q, X has density f_μ, i.e., Q_X is the normal distribution with mean μ and variance σ^2.

2. Show that two Brownian motions $X_{[0,T]}$ and $Y_{[0,T]}$ starting at zero having the same drift parameter μ, but different diffusion coefficients $\sigma_1^2 \neq \sigma_2^2$, have mutually singular distributions. [*Hint:* It suffices to take $\mu = 0$ and consider the law of the iterated logarithm.]
3. (*Probabilistic Construction of Hilbert Transform*) The *Hilbert transform* f_1 of a "rapidly decreasing" C^∞ real-valued function f is defined in harmonic analysis by the requirement that the harmonic extension of $f + if_1$ to the upper half-plane be complex analytic. Here, rapidly decreasing means that f together with all its derivatives vanish at ∞ faster than any power of x. An analytic formula is known to be given in terms of Fourier transform by $\hat{f}_1(\xi) = i\operatorname{sgn}(\xi)\hat{f}(\xi), \xi \in \mathbb{R}$. This exercise concerns an alternative probabilistic representation of f_1. The harmonic extension $u(x, y), x \in \mathbb{R}, y > 0$, of a real-valued function f on $\mathbb{R}$ to the upper half-plane is given by

$$u(x, y) = \mathbb{E}_{(x,y)} f(B_\tau^{(1)})$$

where $B_t = (B_t^{(1)}, B_t^{(2)}), t \geq 0$,is two-dimensional Brownian motion and $\tau := \inf\{t \geq 0 : B_t^{(2)} = 0\}$ is the hitting time of the *boundary* $\mathbb{R}$ to the half-space $\mathbb{R} \times (0, \infty)$.

(a) Show that $u(x, y) = \int_{\mathbb{R}} P_y(x - t)f(t)dt$, where $P_y(x) = \frac{1}{\pi}\frac{y}{y^2+x^2}$ is the so-called *Poisson kernel* (Cauchy density).
(b) Show that the *martingale transform* defined by $V_t = \int_0^t A\nabla u(B_s) \cdot dB_s$, where $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, provides a version of $u_1(B_t)$ where $u_1(x, y)$ is the harmonic extension of the Hilbert transform f_1 of f to the upper half-plane.[*Hint:* Use Itô's lemma and the Cauchy-Riemann equations applied to the analytic function $u(x, y) + iu_1(x, y)$ in the upper half-plane.]
(c) Show that $f_1(x) = \mathbb{E}[V_\tau | B_\tau^{(1)} = x], x \in \mathbb{R}$.

4. Give an example of the exponential martingale $M_{t,\mathbf{x}_0}^0$, $t \geq 0$, as defined in (9.31), where Novikov's condition (9.4) fails. [*Hint:* Use Khaśminskii's criterion of nonexplosion].
5. Let B_t, $t \geq 0$, be a standard one-dimensional Brownian motion with respect to a filtration $\{\mathcal{F}_t : t \geq 0\}$, and let $f(t)$, $t \geq 0$, be a nonanticipative

functional belonging to $\mathcal{L}[0, \infty)$, such that $\int_0^t f^2(s)ds \to \infty$ a.s. as $t \uparrow \infty$. Let $\tau_t := \inf\{s \geq 0 : \int_0^s f^2(u)du = t\}$, $t \geq 0$. Prove that $I_f(\tau_t)$, $t \geq 0$, has the same finite-dimensional distributions as Brownian motion, where $I_f(t) := \int_0^t f(u)dB_u$, $t \geq 0$. [*Hint:* Use characteristic functions] Extend this to k-dimensional stochastic integrals.

6. Let $\mathbf{f}(\cdot)$ be a k-dimensional process belonging to $\mathcal{L}[0, \infty)$. Use the supermartingale property of the exponential martingale M_t defined by (9.1) to prove the *exponential inequality* $P(\sup_{t\geq 0} \int_0^t \mathbf{f}(s) \cdot d\mathbf{B}_s - \frac{\alpha}{2} \int_0^t |\mathbf{f}(s)|^2 ds > \beta) \leq e^{-\alpha\beta}$ $(\alpha > 0, \beta > 0)$.
7. Let $\boldsymbol{\sigma}(t)$, $t \geq 0$, be a $(k \times d)$-matrix valued nonanticipative functional, such that $\int_0^t |\boldsymbol{\sigma}(u)|^2 du < \infty \; \forall t > 0$ and $\int_0^t |\boldsymbol{\sigma}(u)|^2 du \to \infty$ a.s. as $t \to \infty$. Give a direct proof (using characteristic functions) that $I_{\boldsymbol{\sigma}}(t) := \int_0^t \boldsymbol{\sigma}(u)d\mathbf{B}_u$, $t \geq 0$ $(\mathbf{B}_t$, $t \geq 0$, a d-dimensional standard Brownian motion), is a k-dimensional standard Brownian motion if and only if the components $X_t^{(i)} := I_{\boldsymbol{\sigma}_i}(t)$, $1 \leq i \leq k$, ($\boldsymbol{\sigma}_i(\cdot)$ is the ith row of $\boldsymbol{\sigma}(\cdot)$) satisfy $dX_t^{(i)} dX_t^{(j)} = \delta_{ij}dt$ (in the sense of Itô's lemma).
8. Write out a proof of Corollary 9.13.
9. Let $\tau_a = \inf\{t \geq 0 : B_t = a\}$. Define the stochastic process $Y = \{Y_t : t \geq 0\}$ as follows: $Y_t = B_t, t \leq \tau_a$, $Y_t = 2a - B_t, t \geq \tau_a$. (i) Show that Y has the same distribution as B. (ii) Show that $P(B_t \geq x, \tau_a < t) = P(Y_t \geq 2a - x, \tau_a < t) = P(Y_t \geq 2a - x, M_t > a)$, where $M_t = \max\{B_s : 0 \leq s \leq t\}$. (iii) Find the joint distribution of (B_t, M_t). [*Hint:* $B_{\tau_a+s} = B_{\tau_a} + B_{\tau_a+s} - B_{\tau_a}$, and $B_{\tau_a+s} - B_{\tau_a}$, $s \geq 0$, is a standard Brownian motion starting at zero, independent of $\mathcal{F}_{\tau_a}$. In particular, the latter is invariant under reflection about zero. Use this to argue that $\{B_{\tau_a+s:s\geq 0}\}$ has the same distribution as $B_{\tau_a} + (B_{\tau_a} - B_{\tau_a+s})$, $s \geq 0$.]
10. Use the CMG Theorem to derive the joint distribution of $X_t^x := B_t^x + t\mu$ and $\max_{0\leq s\leq t} X_s^x$, where B_t^x, $t \geq 0$, is a standard Brownian motion starting at x. [*Hint:* Use the joint distribution of B_t^x and $\max_{0\leq s\leq t} B_s^x$ (e.g., see Corollary 7.12, p. 84, in Bhattacharya and Waymire (2021)).]
11. Show that the distributions of two Brownian motions with the same diffusion coefficient but different drifts are mutually singular as probabilities on $C[0, \infty)$.[*Hint:* Use the strong law of large numbers].
12. (*Martingale Characterization of Poisson Process*) Suppose that $\{X_t :\geq 0\}$ is a stochastic process with right continuous step function sample paths with positive unit jumps and $X_0 = 0$. Suppose that there is a continuous nondecreasing function $m(t)$ with $m(0) = 0$ such that $M_t := X_t - m(t), t \geq 0$, is a martingale. Show that $\{X_t : t \geq 0\}$ is a (possibly nonhomogeneous) Poisson process.

Chapter 10
Support of Nonsingular Diffusions

This chapter involves the use of the Cameron-Martin-Girsanov theorem to show that the distribution P_x of a nonsingular nonexplosive diffusion is fully supported by $C([0,\infty) : \mathbb{R}^k)$, starting at x. Strong and weak maximum principles for Dirichlet problems associated with the infinitesimal generator are obtained as corollaries.

In this chapter, we use the Cameron-Martin-Girsanov Theorem (Theorem 9.11 or Corollary 9.13) to show that the distribution on $C([0,\infty) : \mathbb{R}^k)$ of every nonexplosive nonsingular diffusion

$$\mathbf{X}_t^{\mathbf{x}_0} = \mathbf{x}_0 + \int_0^t \boldsymbol{\mu}(\mathbf{X}_s^{\mathbf{x}_0})ds + \int_0^t \boldsymbol{\sigma}(\mathbf{X}_s^{\mathbf{x}_0})d\mathbf{B}_s, \quad t \geq 0, \tag{10.1}$$

has the full support $C_{\mathbf{x}_0} := \{\mathbf{f} \in C([0,\infty) : \mathbb{R}^k),\ \mathbf{f}(0) = \mathbf{x}_0\}$.

Theorem 10.1 (Support Theorem) *Suppose $\boldsymbol{\mu}(\cdot)$, $\boldsymbol{\sigma}(\cdot)$ are locally Lipschitzian, $\boldsymbol{\sigma}(\mathbf{x})$ nonsingular for all $\mathbf{x}$, and $\mathbf{X}^{\mathbf{x}_0}$ is nonexplosive. Then the support of the distribution of $\mathbf{X}^{\mathbf{x}_0}$ is $C_{\mathbf{x}_0}$.*

Proof It is enough to prove that for every $T > 0$, $\varepsilon > 0$, and a continuously differentiable $\mathbf{f} \in C_{\mathbf{x}_0}$ with bounded derivative $\mathbf{f}'$, one has

$$P(\max_{0 \leq t \leq T} |\mathbf{X}_t^{\mathbf{x}_0} - \mathbf{f}(t)| \leq \varepsilon) > 0. \tag{10.2}$$

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_10

Let $\mathbf{f}'(t) = \mathbf{c}(t)$, so that $f(t) = \mathbf{x}_0 + \int_0^t \mathbf{c}(u)du$. Then $\mathbf{Z}_t := \mathbf{X}_t^{\mathbf{x}_0} - \mathbf{f}(t)$, $t \geq 0$, is a (nonhomogeneous) diffusion,

$$\mathbf{Z}_t = \int_0^t [\boldsymbol{\mu}(\mathbf{Z}_s + \mathbf{f}(s)) - \mathbf{c}(s)]\, ds + \int_0^t \boldsymbol{\sigma}(\mathbf{Z}_s + \mathbf{f}(s)) d\mathbf{B}_s. \tag{10.3}$$

Then (10.2) is equivalent to

$$P(\max_{0 \leq t \leq T} |\mathbf{Z}_t| \leq \varepsilon) > 0. \tag{10.4}$$

By Theorem 9.9, this is equivalent to proving

$$P(\max_{0 \leq t \leq T} |\mathbf{Y}_t| \leq \varepsilon) > 0, \tag{10.5}$$

where $\mathbf{Y}$ is governed by the equation

$$\mathbf{Y}_t = \int_0^t -M\mathbf{Y}_s ds + \int_0^t \boldsymbol{\sigma}_{\mathbf{f}}(\mathbf{Y}_s) d\mathbf{B}_s, \tag{10.6}$$

$\boldsymbol{\sigma}_{\mathbf{f}}(s, y) := \boldsymbol{\sigma}(y + \mathbf{f}(s))$, and M is a (large) positive number to be specified later. Note that for a fixed $\mathbf{f}$ and $\varepsilon > 0$, the probability in (10.5) depends only on the values of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ in a compact neighborhood of $\mathbf{x}_0$. Hence, we may assume, without loss of generality, that $\boldsymbol{\mu}_{\mathbf{f}}(s, y) := \boldsymbol{\mu}(y + \mathbf{f}(s))$, $\boldsymbol{\sigma}_{\mathbf{f}}$, $\boldsymbol{\sigma}_{\mathbf{f}}^{-1}$ are bounded and Lipschitzian (to satisfy the hypothesis for $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ in Corollary 9.10). Now (10.5) is implied by

$$\mathbb{E}\tau \geq T, \qquad \tau := \inf\{t \geq 0 : |\mathbf{Y}_t| = \varepsilon\}. \tag{10.7}$$

To derive the last inequality (for a sufficiently large M), define

$$\mathbf{a}(s, \mathbf{y}) := \boldsymbol{\sigma}_{\mathbf{f}}(s, \mathbf{y})\boldsymbol{\sigma}'_{\mathbf{f}}(s, y), \qquad d(s, \mathbf{y}) := \sum_{i,j} a_{ij}(s, \mathbf{y}) y^{(i)} y^{(j)} / |\mathbf{y}|^2,$$

$$B(s, \mathbf{y}) := \sum_i a_{ii}(s, \mathbf{y}), \qquad C(\mathbf{y}) := -2M|\mathbf{y}|^2. \tag{10.8}$$

Then applying the generator $A_s := \frac{1}{2} \sum a_{ij}(s, \mathbf{y}) \partial^2 / \partial y^{(i)} \partial y^{(j)} - M \sum y^{(i)} \partial / \partial y^{(i)}$ to a radial function $\varphi(\mathbf{y}) = F(|\mathbf{y}|)$, with $F(\cdot)$ twice differentiable on $[0, \varepsilon]$,

$$\begin{aligned} A_s \varphi(\mathbf{y}) &= \frac{1}{2} a(s, \mathbf{y}) F''(|\mathbf{y}|) + \frac{1}{2} \left(\frac{B(s, \mathbf{y}) + C(\mathbf{y}) - a(s, \mathbf{y})}{|\mathbf{y}|}\right) F'(|\mathbf{y}|) \\ &= \frac{1}{2} a(s, \mathbf{y}) [F''(|\mathbf{y}|) + \{\frac{B(s, \mathbf{y}) + C(\mathbf{y})}{a(s, \mathbf{y})} - 1\} \frac{1}{|\mathbf{y}|} F'(|\mathbf{y}|)] \end{aligned} \tag{10.9}$$

We would like to find a function F on $[0,\varepsilon]$ such that $F(0) > T$, $F(\varepsilon) = 0$, $F'(r) < 0$ and $A_s\varphi(\mathbf{y}) \geq -1$ for $0 \leq s \leq T$, $|\mathbf{y}| \equiv r \leq \varepsilon$. For then Itô's lemma and optional stopping would yield $\mathbb{E}\varphi(\mathbf{Y}_\tau) - \varphi(\mathbf{0}) \geq \mathbb{E}(-\tau)$, or, $\mathbb{E}\tau \geq F(0) > T$. One such function is

$$F(r) := \frac{2}{\lambda}\int_r^\varepsilon \Big(\int_0^u \exp\{-\int_v^u \frac{\beta(v')}{v'}\,dv'\}dv\Big)du,$$

$$\beta(u) := \frac{k\lambda}{\Lambda} - \frac{2Mu^2}{\lambda} - 1, \tag{10.10}$$

where $\lambda > 0$ is a lower bound and $\Lambda < \infty$ is an upper bound of the eigenvalues of $\mathbf{a}(s,\mathbf{y})$ for $0 \leq s \leq T$, $|\mathbf{y}| \leq \varepsilon$.

Note that $\{B(s,\mathbf{y}) + C(\mathbf{y})\}/d(s,\mathbf{y}) - 1 \geq \beta(|\mathbf{y}|)$ for $0 \leq s \leq T$, $|\mathbf{y}| \leq \varepsilon$. Also

$$F'(r) = -\frac{2}{\lambda}\int_0^r \exp\{-\int_v^r \frac{\beta(v')}{v'}\,dv'\}dv < 0,$$

$$F''(r) = -\frac{2}{\lambda}\Big[1 + \int_0^r \Big(-\frac{\beta(r)}{r}\Big)\exp\{-\int_v^r \frac{\beta(v')}{v'}\,dv'\}dv\Big]$$

$$= -\frac{2}{\lambda} - \frac{\beta(r)}{r}F'(r),$$

so that

$$F''(|\mathbf{y}|) + \Big\{\frac{B(s,\mathbf{y}) + C(\mathbf{y})}{d(s,\mathbf{y})} - 1\Big\}\frac{1}{|\mathbf{y}|}F'(|\mathbf{y}|) \geq -\frac{2}{\lambda},$$
$$0 \leq s \leq T,\ \ |\mathbf{y}| \leq \varepsilon. \tag{10.11}$$

Using (10.11) in (10.9), we get

$$A_s\varphi(\mathbf{y}) \geq -1 \qquad 0 \leq s \leq T,\ 0 < |\mathbf{y}| \leq \varepsilon. \tag{10.12}$$

Applying Itô's lemma (and optional stopping rule) to $\varphi(\mathbf{Y}_t)$ (see Exercise 1) to get $\mathbb{E}\tau \geq F(0)$. From (10.10), it is clear that $F(0) \to \infty$ as $M \to \infty$. Hence we can choose M such that $\mathbb{E}\tau \geq T$. ■

Corollary 10.2 (The Strong Maximum Principle) *Let G be a connected open set, and $A = \frac{1}{2}\sum a_{ij}(x)\partial^2/\partial x^{(i)}\partial x^{(j)} + \sum b_i(x)\partial/\partial x^{(i)}$ be the generator of a diffusion such that $a^{-1}(\cdot) \equiv (\boldsymbol{\sigma}(\cdot)\boldsymbol{\sigma}(\cdot)')^{-1}$ is bounded. If $u(\cdot)$ is A–harmonic, i.e., $Au(\mathbf{x}) = 0\ \forall\,\mathbf{x} \in G$, then u cannot attain its maximum or minimum in G unless u is constant on G.*

Proof Let $\mathbf{x} \in G$ be such that one has $u(\mathbf{x}) \geq u(\mathbf{y})\ \forall\,\mathbf{y} \in G$. Let $\delta > 0$ be such that $\overline{B(\mathbf{x}:\delta)} := \{\mathbf{y} : |\mathbf{y} - \mathbf{x}| \leq \delta\} \subseteq G$. Consider the diffusion $\mathbf{X}^{\mathbf{x}}$ with generator A, starting at $\mathbf{x}$. Then, by the strong Markov property, one has (Exercise 2)

$$u(\mathbf{x}) = \mathbb{E}\left(u(\mathbf{X}_\tau^{\mathbf{x}})\right) \tag{10.13}$$

where $\tau := \inf\{t \geq 0 : |\mathbf{X}_t^{\mathbf{x}} - \mathbf{x}| = \delta\}$. If $\pi(\mathbf{x}, d\mathbf{y})$ denotes the distribution of $\mathbf{X}_\tau^{\mathbf{x}}$ (on $\{\mathbf{y} : |\mathbf{y} - \mathbf{x}| = \delta\}$), then (10.13) may be expressed as

$$u(\mathbf{x}) = \int_{\{\mathbf{y}:|\mathbf{y}-\mathbf{x}|=\mathbf{d}\}} u(\mathbf{y})\pi(\mathbf{x}, d\mathbf{y}). \tag{10.14}$$

It follows easily from the Support Theorem 10.1 that the support of $\pi(\mathbf{x}, d\mathbf{y})$ is (all of) $\{\mathbf{y} : |\mathbf{y} - \mathbf{x}| = \delta\}$ (Exercise 3). Since u is continuous (on $\{\mathbf{y} : |\mathbf{y} - \mathbf{x}| = \delta\}$) and $u(\mathbf{y}) \leq u(\mathbf{x})\ \forall\, \mathbf{y}$, one must have $u(\mathbf{y}) = u(\mathbf{x})\ \forall\, \mathbf{y} \in \{\mathbf{y} : |\mathbf{y} - \mathbf{x}| = \delta\}$ in view of (10.14).

Since the same is true for all $\delta' \in (0, \delta]$, $u(\mathbf{y}) = u(\mathbf{x})\ \forall\, \mathbf{y} \in B(\mathbf{x} : \delta)$. This shows that the set B of points $\mathbf{x} \in G$ such that $u(\mathbf{x}) = \sup_{\mathbf{z}\in G} u(\mathbf{z})$ is an open set. But B is also a closed subset of G, since $B = \{\mathbf{x} \in G : u(\mathbf{x}) = M\}$, where $M = \sup_{\mathbf{z}\in G} u(\mathbf{z})$. Hence B is either empty or $B = G$. ∎

Among the important applications of maximum principles are those to uniqueness problems for important classes of partial differential equations (also see Chapter 16).

Corollary 10.3 (The Weak Maximum Principle) *Let G be a bounded open set, and let A satisfy the hypothesis of Corollary 10.2. If f is a continuous function on ∂G, and g is a continuous function on G, and if u_1, u_2 are both solutions to the Dirichlet problem*

$$Au_j(\mathbf{x}) = g(\mathbf{x}),\ \ (\mathbf{x} \in G), \quad \lim_{\mathbf{x}\to\mathbf{b}} u_j(\mathbf{x}) = f(\mathbf{b})\ \ (\mathbf{b} \in \partial G),\ \ (j = 1, 2), \tag{10.15}$$

then $u_1 \equiv u_2$ (on $\overline{G}$).

Proof Without essential loss of generality, assume G is connected (else, apply the following argument to each connected component of G). Since $v := u_1 - u_2$ satisfies $Av(\mathbf{x}) = 0$ for $\mathbf{x} \in G$, $v(\mathbf{b}) = 0$ for $\mathbf{b} \in \partial G$, and v is continuous on $\overline{G}$, $\mathbf{v}$ attains its minimum as well as maximum on ∂G, which are both zero. Therefore $v \equiv 0$ on $\overline{G}$. ∎

Exercises

1. (i) For $k = 1$, show that $\varphi(y)$ in Theorem 10.1 is twice continuously differentiable on $(-\varepsilon, \varepsilon)$, and may be extended to a bounded function with bounded first and second derivatives on all of $\mathbb{R}$. Then justify the use of Itô's lemma and the optional stopping rule to derive the desired results $\mathbb{E}\tau \geq T$.
   (ii) For $k > 1$, and arbitrary $0 < \delta < \varepsilon$, show that $\varphi(\mathbf{y})$ is twice continuously differentiable on $\delta < |\mathbf{y}| < \varepsilon$, which can be extended to a bounded

function φ_δ with bounded first and second order derivatives on $\mathbb{R}^k$. [*Hint:* Derive expressions for derivatives of a radial function, or look ahead to the display in the proof of Lemma 8 in the next chapter.] Starting at $\mathbf{y}_0$ such that $\delta < |\mathbf{y}_0| < \varepsilon$, using Itô's lemma to this extended function, show that $\mathbb{E}\tau_\delta^{\mathbf{y}_0} \wedge \tau_\varepsilon^{\mathbf{y}_0} \wedge t \geq F(|\mathbf{y}_0|) - \mathbb{E}F(|\mathbf{Y}_{\tau_\delta^{\mathbf{y}_0} \wedge \tau_\varepsilon^{\mathbf{y}_0} \wedge t}|)$ for all $t \geq 0$, where $\tau_\theta^{\mathbf{y}_0} = \inf\{t \geq 0 : |\mathbf{Y}_t^{\mathbf{y}_0}| = \theta\}$, $\theta > 0$. Then let $t \uparrow \infty$, and then let $\delta \downarrow 0$ to get $\mathbb{E}\tau_{0+}^{\mathbf{y}_0} \wedge \tau_\varepsilon^{\mathbf{y}_0} \geq F(|\mathbf{y}_0|) - \mathbb{E}F(|\mathbf{Y}_{\tau_{0+}^{\mathbf{y}_0} \wedge \tau_\varepsilon^{\mathbf{y}_0}}|)$, where $\tau_{0+}^{\mathbf{y}_0} := \lim_{\delta \downarrow 0} \tau_\delta^{\mathbf{y}_0}$. Prove that $\tau_{0+}^{\mathbf{y}_0} = \infty$ a.s., and first derive $\mathbb{E}\tau_\varepsilon^{\mathbf{y}_0} \geq F(|\mathbf{y}_0|)$, and then $\mathbb{E}\tau_\varepsilon^{\mathbf{0}} \geq F(0)$.

2. Derive (10.13).
3. Let τ be the hitting time of the sphere $\{\mathbf{y} : |\mathbf{y} - \mathbf{x}| = \delta\}$. Show that the support of the distribution $\pi(\mathbf{x}, d\mathbf{y})$ of $\mathbf{X}_\tau^{\mathbf{x}}$ is full under the hypothesis of Corollary 10.2.

Chapter 11
Transience and Recurrence of Multidimensional Diffusions

> In this chapter criteria are developed for the transience, null recurrence, and positive recurrence of multidimensional diffusions generated by a second order elliptic operator, denoted by L, with the diffusion matrix $((a_{ij}(\cdot)))_{1\le i,j\le k}$, non-singular and continuous, and drift vector $b(\cdot)$, measurable and bounded on compact subsets of $\mathbb{R}^k$.

In this chapter we analyze[1] the long time behavior of diffusions generated by the elliptic operator

$$L = \frac{1}{2}\sum_{i,j=1}^{k} a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{j=1}^{k} b_j(x)\frac{\partial}{\partial x_j} \tag{11.1}$$

under the following conditions assumed throughout this chapter. Criteria[2] were announced by Khaśminskii (1960) under additional smoothness conditions.

Remark 11.1 The new notation L is being used here, in place of the earlier notation A in preceding chapters for elliptic operators generating a diffusion, in order to indicate that the smoothness assumptions for A are relaxed here.

---

[1] The generality in the approach presented in this chapter follows Bhattacharya (1978).

[2] Unaware of Khaśminskii's (1960) announcement, Friedman (1973) derived analagous criteria for recurrence and transience under additional growth and smoothness conditions. Alternative derivations of Khaśminskii's announced results under additional smoothness conditions may be found in Pinsky and Dieck (1995).

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_11

Condition (A): The matrix $((a_{ij}(x)))$ is real-symmetric and positive definite for each $x \in \mathbb{R}^k$, and the functions $a_{ij}(x)$, $1 \le i, j \le k$, are continuous. The functions $b_j(x)$, $1 \le j \le k$, are real-valued, Borel measurable and bounded on compacts.

A key role is played by the following property of diffusions to be considered here.

Definition 11.1 (Strong Feller Property) A Markov process $X = \{X(t) : t \ge 0\}$ starting at $x \in \mathbb{R}^k$ under P_x is said to have the strong Feller property under P_x if for every bounded measurable function $f : \mathbb{R}^k \to \mathbb{R}$, the function $x \to \mathbb{E}_x f(X(t))$ is continuous on $\mathbb{R}^k$ for each fixed $t > 0$.

The following extension of the construction of diffusions to possibly non-smooth coefficients on $\mathbb{R}^k$, $k \ge 2$, is due to Stroock and Varadhan (1979). For the case of Lipschitz coefficients this is a consequence of Itô's lemma (see Corollary 8.5), and Theorem 10.1.

Theorem 11.1 *In addition to the Condition (A), suppose that a_{ij}, b_j are bounded on $\mathbb{R}^k$. Then for each $x \in \mathbb{R}^k$ there is a unique probability measure P_x on $\Omega' = C([0,\infty) : \mathbb{R}^k)$ with Borel σ-algebra for the topology of uniform convergence on compact subsets of $[0,\infty)$. Then,*

(i) $P_x(X(0) = x) = 1, \ x \in \mathbb{R}^k$
(ii) *For every bounded $f : \mathbb{R}^k \to \mathbb{R}$, having bounded and continuous first and second order derivatives, the process*

$$M(t) = f(X(t)) - \int_0^t Lf(X(s))ds, \quad t \ge 0,$$

is a martingale under P_x with respect to the filtration $\mathcal{F}_t = \sigma(X(u) : u \le t)$, $t \ge 0$.

Furthermore, (a) X is a strong Markov and strong Feller process, and (b) The support of P_x is $\{\omega \in C([0,\infty) : \mathbb{R}^k) : \omega(0) = x\}$.

Remark 11.2 It may be noted that Theorem 11.1 holds more generally under condition A if explosion does not occur (see Definition 11.4). A sufficient condition is provided in the next chapter.

The first step to relax the conditions for this construction to those of Condition (A) is to consider continuous truncations of the coefficients of L outside a large ball centered at the origin. Specifically, for each integer $N \ge 1$, define truncated coefficients of L as follows:

(i) $a_{ij,N}(x) = a_{ij}(x), \quad b_{j,N}(x) = b_j(x)$, for $|x| \le N$, $1 \le i, j \le k$.
(ii) $a_{ij,N}(x) = a_{ij}(x_0), \quad b_{j,N}(x) = b_{j,N}(x_0)$ for $x = cx_0$ such that $|x_0| = N, c > 1$.

In addition, define

$$L_N = \frac{1}{2}\sum_{i,j=1}^{k} a_{ij,N}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{j=1}^{k} b_{j,N}(x)\frac{\partial}{\partial x_j}, \quad N = 1, 2, \ldots .$$

Definition 11.2 The diffusion $P_{x,N}, x \in \mathbb{R}^k$, or the coordinate processes on $C([0,\infty) : \mathbb{R}^k)$, associated with distributions $P_{x,N}, x \in \mathbb{R}^k$, associated with the operator L_N in accordance with Theorem 11.1, will be referred to as the diffusion with generator L_N.

Under Condition (A) it is a somewhat technical matter that one can construct a canonical model of the diffusion process on the enlarged state space $\mathbb{R}^k \cup \{\infty\}$ obtained by adjoining a fictitious state, denoted ∞, to $\mathbb{R}^k$ under the one-point compactification topology (see Stroock and Varadhan (1979).) Under the additional condition that the coefficients of L are Lipschitz, the result follows from Corollary 8.2 and Remark 8.1.

Theorem 11.2 *Assume Condition (A) holds. Then for each $x \in \mathbb{R}^k$ there is a probability measure P_x on the Borel σ-field of $\Omega = C([0,\infty) : \mathbb{R}^k \cup \{\infty\})$ for the one-point compactification of $\mathbb{R}^k \cup \{\infty\}$, and the corresponding topology of uniform convergence on compact subsets of $[0,\infty)$ on Ω such that the coordinate projection process $X = \{X(t) : t \geq 0\}$ on Ω is a strong Markov process under $P_x, x \in \mathbb{R}^k \cup \{\infty\}$, and for each twice continuously differentiable function f,*

$$M(t) = f(X(t)) - \int_0^t Lf(X(s))ds, \quad t \geq 0,$$

is a martingale under each $P_x, x \in \mathbb{R}^k \cup \{\infty\}$.

We will continue to represent the coordinate-projection process with state space $\mathbb{R}^k \cup \{\infty\}$ with the same notation, X and P_x. The coordinate-projection processes associated with $L_N, N \geq 1$, through Theorem 11.1 will continue to be denoted by $P_{x,N}, x \in \mathbb{R}^k$. Also $\mathcal{F}_t$ will denote the filtration generated by $\sigma(X(s) : 0 \leq s \leq t)$, $t \geq 0$, and $\mathcal{F}_\tau$ denotes the pre-τ sigmafield if τ is a stopping time (see BCPT[3] p. 59).

Recall that for an open set $U \subset \mathbb{R}^k$, and $x \in U$, the *escape time*

$$\eta_U(x) = \inf\{t \geq 0 : X(t) \notin U\} = \inf\{t \geq 0 : X(t) \in \partial U\}, \tag{11.2}$$

is a stopping time.

Definition 11.3 Let L denote the elliptic operator (11.1) under the assumptions of Condition (A), and let $X = \{X(t) : t \geq 0\}$ be the coordinate-projection process having distributions $P_x, x \in \mathbb{R}^k$. Let

[3] Throughout, BCPT refers to Bhattacharya and Waymire (2016), A Basic Course in Probability Theory.

$$B(x : r) = \{y \in \mathbb{R}^k : |y - x| < r\}.$$

(i) A point $x \in \mathbb{R}^k$ is said to be recurrent for the diffusion if for any $\varepsilon > 0$,

$$P_x(X(t) \in B(x : \varepsilon) \text{ for a sequence of } t \uparrow \infty) = 1.$$

(ii) A point $x \in \mathbb{R}^k$ is said to be transient if

$$P_x(|X(t)| \to \infty \text{ as } t \to \infty) = 1.$$

If all points are recurrent, or if all points are transient, then the diffusion is said to be recurrent, or to be transient, respectively.

The following lemma follows from the main theorem (see Stroock and Varadhan (1979), Lemma 9.2.2, page 234), and is stated without proof. For elliptic operators with Hölder continuous coefficients, it is a standard result in the theory of partial differential equations (see Friedman (1964)) that the transition probability has a density $p(t; x, y)$ which is differentiable in both x, y, for all $t > 0$.

Lemma 1 *Assume Condition (A). For each Borel measurable set $B \subset \mathbb{R}^k$, the map $(t, x) \to p(t; x, B)$ is continuous on $(0, \infty) \times \mathbb{R}^k$. In particular, for all $(t, x) \in (0, \infty) \times \mathbb{R}^k$, $p(t; x, dy)$ is absolutely continuous with respect to Lebesgue measure with a density $p(t; x, y) \geq 0$, in the sense that $p(t; x, B) = \int_B p(t; x, y)dy$, for all $B \in \mathcal{S}$.*

Lemma 2 *Let (S, ρ) be a metric space with Borel σ-field $\mathcal{S}$. Let $(x, t) \to p(t; x, B)$ be continuous for each $B \in \mathcal{S}$. Consider a Markov process $\{X(t) : t \geq 0\}$ on a probability space $(\Omega, \mathcal{F}, P_x)$, $x \in S$, having transition probability $p(t; x, dy)$. Then for every bounded measurable function g on Ω, the function $x \to \mathbb{E}_x g$ is continuous on S.*

Proof Define the filtration $\mathcal{F}_t = \sigma(X(s), 0 \leq s \leq t), t \geq 0$. Suppose g is a finite dimensional measurable function on Ω, say,

$$g(X) = g(X(0), X(t_1), \ldots, X(t_k)), \quad 0 < t_1 < \cdots < t_k.$$

Then,

$$\begin{aligned} \mathbb{E}_x g &= \mathbb{E}_x[\mathbb{E}(g(X(0), X(t_1), \ldots, X(t_k))|\mathcal{F}_{t_{k-1}})] \\ &= \mathbb{E}_x g_{k-1}(X(0), X(t_1), \ldots, X(t_{k-1})), \end{aligned} \tag{11.3}$$

where $g_{k-1}(x, x_1, \ldots, x_{k-1}) = \int_S g(x, x_1, \ldots, x_{k-1}, x_k)p(t_k - t_{k-1}; x_{k-1}, dx_k)$. Continuing in this manner, $g_1(x, x_1) = \int_S g(x, x_1, x_2)p(t_2 - t_1; x_1, dx_2)$, and therefore, $\mathbb{E}_x g = \int_S g_0(x, x_1)p(t_1; x, dx_1)$. By assumption, the last integral is

continuous, since a bounded measurable function can be approximated uniformly by simple functions.[4] Restricting to indicator functions $g = \mathbf{1}_B$ of measurable sets, it follows that the class $\mathcal{C}$ of sets such that $x \to P_x(B)$ is continuous contains the field of all finite dimensional sets. But if $B_n \uparrow B$ then $P_x(B_n) \uparrow P_x(B)$ for B_n a finite dimensional sequence. Therefore, $x \to P_x(B)$ is lower semicontinuous for all sets B belonging to the σ-field generated by the field of finite dimensional sets. Turning to complements, it follows that $x \to P_x(B^c) = 1 - P_x(B)$ is lower semicontinuous. That is $x \to P_x(B)$ is upper semicontinuous. Hence $x \to P_x(B)$ is continuous for all B in $\mathcal{C}$, which is, therefore, the σ-field $\mathcal{F}$. An entirely analogous argument holds for bounded measurable functions. ■

Definition 11.4 (Explosion Time) Assume Condition (A). The explosion time for X is defined by

$$\zeta = \lim_{N \uparrow \infty} \eta_{B(0:N)},$$

where $B(0:N) = \{x \in \mathbb{R}^k : |x| < N\}$, and $\eta_{B(0:N)}$ is the escape time from the ball $B(0:N)$. For $x \in \mathbb{R}^k$, the probability measure P_x is said to be conservative if $P_x(\zeta = \infty) = 1$.

Definition 11.5 Assume Condition (A). A Borel measurable function $f : \mathbb{R}^k \to \mathbb{R}$ is said to be L-harmonic on an open set G if it is bounded on compacts, and for all $x \in G$,

$$f(x) = \mathbb{E}_x f(X(\eta_U)),$$

for every neighborhood U of x having compact closure $\overline{U}$ in G.

Under additional smoothness of L, the following lemma follows from Corollary 8.3 and Corollary 10.3

Lemma 3 *Assume Condition (A) holds.*

a. *Every L-harmonic function on an open set $G \subset \mathbb{R}^k$ is continuous on G.*
b. *(Maximum Principle) Let f be a non-negative L-harmonic function on a connected open set $G \subset \mathbb{R}^k$. Then f is either strictly positive on G or $f \equiv 0$ on G.*

Proof To prove part (a), let f be a L-harmonic function on an open set $G \subset \mathbb{R}^k$. For $x \in U$, an open subset of G whose closure $\overline{U}$ in G is compact, one has

$$f(x) = \mathbb{E}_x f(X(\eta_U)) = \mathbb{E}_{x,N} f(X(\eta_U)),$$

[4] See BCPT, p. 221, (2.4).

for $B(0:N) \supset U$. This makes f continuous on G, by Lemma 2. For part (b), suppose that $f \geq 0$ on G and $f(x_0) = 0, x_0 \in G$, and let B denote the open ball centered at x_0 of radius $\varepsilon > 0$, sufficiently small that $B \subset G$. Then

$$0 = f(x_0) = \mathbb{E}_{x_0} f(X(\eta_B)) = \int_{\partial B} f(y) p(x_0, dy),$$

where $p(x_0, dy)$ is the distribution of $X(\eta_B)$ under P_{x_0}, and therefore under $P_{x_0,N}$ for $N > |x_0| + \varepsilon$, as well. By Theorem 11.1, the support of $p(x_0, dy)$ is ∂B (Exercise 1). Since $f \geq 0$ and continuous, it follows that $f = 0$ on ∂B. Thus, $f = 0$ on G. ■

Lemma 4 *Assume Condition (A) holds.*

a. *If $U \subset \mathbb{R}^k$ is a non-empty, open subset of $\mathbb{R}^k$, such that there is a non-empty open connected set V, $\overline{V} \cap \overline{U} = \emptyset$, then $x \to P_x(\eta_U < \infty)$ is positive and continuous on U.*

b. *If $U_1, U_2 \subset \mathbb{R}^k$ are two non-empty, open subsets of $\mathbb{R}^k$, such that $\overline{U}_1 \cap \overline{U}_2 = \emptyset$, and $\overline{U}_2^c = \mathbb{R}^k \backslash \overline{U}_2$ is connected, then $x \to P_x(\eta_{\overline{U}_1^c} < \eta_{\overline{U}_2^c})$ is positive and continuous on $\overline{U}_1^c \cap \overline{U}_2^c$, where $\eta_{\overline{U}_i^c}$ is the escape time of the open set $\overline{U}_i^c$, $i = 1, 2$.*

Proof It follows from the strong Markov property that both functions are L-harmonic and, therefore, continuous. To prove positivity in (b), and therefore positivity in (a), let $x \in \overline{U}_1^c \cap \overline{U}_2^c$. Let $B \subset U_1$ be an open ball and B_1 a bounded open set such that $x \in B_1$, $B \subset B_1 \subset \{x \in \mathbb{R}^{\cdot} |x| < N\}$, $B_1 \cap \overline{U}_2 = \emptyset$. Then,

$$P_x(\eta_{\overline{U}_1^c} < \eta_{\overline{U}_2^c}) \geq P_x(\eta_{\overline{B}^c} < \eta_{\overline{B}_1^c}) = P_{x,N}(\eta_{\overline{B}^c} < \eta_{\overline{B}_1^c}) > 0.$$

The final expression is positive since

$$P_{x,N}(\{\omega \in C([0,\infty) : \mathbb{R}^k) : \omega(0) = x\}) = 1.$$

■

Lemma 5 *Assume Condition (A) holds. If P_{x_0} is conservative for some $x_0 \in \mathbb{R}^k$, then P_x is conservative for all $x \in \mathbb{R}^k$, and the process $\{X(t) : t \geq 0\}$ has the strong Feller property.*

Proof Since $x \to P_x(\zeta < \infty)$ is harmonic, the first assertion follows. The strong Feller property follows from Lemma 2. ■

Lemma 6 *Assume Condition (A) holds. If U is a bounded open subset of $\mathbb{R}^k$ with escape time η_U, then for all $t > 0$, $x \in \mathbb{R}^k$,*

$$\mathbb{E}_x \eta_U \leq \frac{t}{1 - \sup_x P_x(\eta_U > t)}. \tag{11.4}$$

Proof For fixed but arbitrary $t > 0$, let $q = \sup_x P_x(\eta_U > t)$, and $A_t = [\eta_U > t]$. Then, by the strong Markov property, for $s \geq 0$,

$$P_x(A_{t+s}) = \mathbb{E}_x\{\mathbf{1}_{A_s} P_{X(s)}(A_t)\} \leq q P_x(A_s).$$

Iterating, one has

$$P_x(A_{nt}) \leq q^n, \quad n = 1, 2, \ldots.$$

Thus,

$$\begin{aligned}
\mathbb{E}_x \eta_U &= \sum_{n=0}^{\infty} \mathbb{E}_x \mathbf{1}_{[nt < \eta_U \leq (n+1)t]} \eta_U \\
&\leq \sum_{n=0}^{\infty} (n+1)t P_x(nt < \eta_U \leq (n+1)t) \\
&= \sum_{n=0}^{\infty} (n+1)t\{P_x(A_{nt})) - P_x(A_{(n+1)t})\} \\
&= \sum_{n=0}^{\infty} t P_x(A_{nt}) \leq t \sum_{n=0}^{\infty} q^n = \frac{t}{1-q}.
\end{aligned}$$

■

Remark 11.3 As an aside, it is interesting to notice that this argument extends to the following bound, for any integer $k \geq 1$,

$$\mathbb{E}_x \mathbf{1}_{[\eta_U \geq kt]} \leq t \sum_{n=k}^{\infty} q^n = \frac{q^k}{1-q},$$

where, for arbitrary but fixed $t > 0$, $q = \sup_x P_x(\eta_U > t)$.

Lemma 7 *Assume Condition (A) holds. If U is a bounded open subset of $\mathbb{R}^k$, then*

$$\sup_{x \in U} \mathbb{E}\eta_U < \infty.$$

Proof Let N be sufficiently large so that $B(0:N) \supset \overline{U}$. Then $\mathbb{E}_x \eta_U = \mathbb{E}_{x,N} \eta_U$ for $x \in U$. Fix $t_0 > 0$. Since $P_{x,N}(\omega \in C([0,\infty):\mathbb{R}^k):\omega(0) = x) = 1$, one has

$$\sup_{x \in \overline{U}} P_{x,N}(\eta_U > t_0) \leq P_{x,N}(|X(t_0)| < N) < 1, \ \forall\, x \in \overline{U}.$$

The assertion now follows by applying Lemma 6. ■

Proposition 11.3 *Assume Condition (A) holds. The following are equivalent.*

a. *The diffusion is recurrent.*
b. $P_x(X(t) \in U$ *for some* $t \geq 0) = 1$ *for all* $x \in \mathbb{R}^k$, *nonempty open sets* $U \subset \mathbb{R}^k$.
c. *There is a compact set* $K \subset \mathbb{R}^k$ *such that* $P_x(X(t) \in K$ *for some* $t \geq 0) = 1$ *for all* $x \in \mathbb{R}^k$.
d. $P_x(X(t) \in U$ *for a sequence of* $t \uparrow \infty) = 1$ *for all* $x \in \mathbb{R}^k$, *nonempty open sets* $U \subset \mathbb{R}^k$.
e. *For each* $x \in \mathbb{R}^k$, *there is a point* $z \in \mathbb{R}^k$, *a pair of numbers* $0 < r_0 < r_1$, *and a point* $y \in \partial B(z : r_1) = \{w : |w - z| = r_1\}$ *such that* $P_y(\eta_{\overline{B}^c(x:r_0)} < \infty) = 1$.

Proof The implications (b)⇒(c), (b)⇒(e), and (d)⇒(a) are obvious. To prove (a)⇒(b), let $x \in \mathbb{R}^k$, and let U be a nonempty open set such that $x \notin U$. Let $B \subset U$ be an open ball, and choose $\varepsilon > 0$ such that $\overline{B(x : \varepsilon)} \cap \overline{B} = \emptyset$. Let $U_1 \supset \overline{B(x : \varepsilon)} \cup \overline{B}$ be a bounded open set. Define $\eta_1 = \eta_{U_1}$, $\eta_{2i} = \inf\{t > \eta_{2i-1} : X(t) \in \partial B(x : \varepsilon)\}$, $\eta_{2i+1} = \inf\{t > \eta_{2i} : X(t) \in \partial U_1\}$, $i = 1, 2, \ldots$. By Lemma 7, and recurrence of x, the η_i's are P_x-a.s. finite stopping times. Consider the events,

$$A_0 = [X(t) \in \overline{B} \text{ for some } t \in [0, \eta_1)],\ A_i = [X(t) \in \overline{B} \text{ for some } t \in [\eta_{2i-1}, \eta_{2i})],\ i \geq 2.$$

Since $y \to P_y(\eta_{\overline{B}^c} < \eta_{\overline{B(x:\varepsilon)}^c}) > 0$ is positive and continuous on $\overline{B}^c \cap \overline{B(x : \varepsilon)}^c$ by Lemma 4(b),

$$\delta = \inf_{y \in \partial U_1} P_y(\eta_{\overline{B}^c} < \eta_{\overline{B(x:\varepsilon)}^c}) > 0.$$

Using the strong Markov property and induction on n, one obtains $P_x(\cap_{i=0}^n A_i^c) \leq (1 - \delta)^n$. Thus,

$$P_x(X(t) \in U \text{ for no } t \geq 0) \leq P_x(X(t) \in \overline{B} \text{for no } t \geq 0) \leq \lim_n P_x\left(\cap_{i=0}^n A_i^c\right) = 0.$$

Next, to prove (b)⇒(d), let $x \in \mathbb{R}^k$, U a nonempty open set, B an open ball, and $\varepsilon > 0$ such that $\overline{B} \cap \overline{B(x : \varepsilon)} = \emptyset$ and $\overline{B} \subset U$. Define

$$\theta_1 = \inf\{t \geq 0 : X(t) \in \partial B(x : \varepsilon)\}, \theta_{2i} = \inf\{t \geq \theta_{2i-1} : X(t) \in \partial B\},$$

and

$$\theta_{2i+1} = \inf\{t \geq \theta_{2i} : X(t) \in \partial B(x : \varepsilon)\},\ i \geq 1.$$

By (b) and the strong Markov property, the θ_i's are P_x-a.s. finite. Also $\theta_i \uparrow \infty$ P_x-a.s. as $i \uparrow \infty$; for otherwise, with P_x-positive probability the sequences $\{X(\theta_{2i-1})\}_{i=1}^\infty$ and $\{X(\theta_{2i})\}_{i=1}^\infty$ converge to a common limit, which is impossible since $\partial B(x : \varepsilon)$ and ∂B are disjoint.

To prove (c)⇒(b), Let K be as in (c), B an arbitrary open ball, and $x \in \mathbb{R}^k$. Let U be an open ball containing $\overline{B} \cup K$. Define

$$\eta'_1 = \eta_{K^c},$$
$$\eta'_{2i} = \inf\{t \geq \eta'_{2i-1} : X(t) \in \partial U\},$$
$$\eta'_{2i+1} = \inf\{t \geq \eta'_{2i} : X(t) \in K\},\ i \geq 1.$$

By (c), the strong Markov property and Lemma 7, η_i, $i \geq 1$, are P_x-a.s. finite. Now proceed as in the proof that (a)⇒(b). Specifically, define the A_i's by replacing η_i by η'_i, $i \geq 1$, respectively, and let $\delta + \inf_{y \in K} P_y(A_1)$. then $\delta > 0$ by Lemma 4(b), and

$$P_x(X(t) \in \overline{B} \text{ for no } t \geq 0) \leq P_x(\cap_{i=1}^n A_i^c) \leq (1-\delta)^n, \quad n = 1, 2 \ldots.$$

To prove (e)⇒(c), take $K = \overline{B(z : r_0)}$ and use Lemma 4(a), and the maximum principle of Lemma 3(b). ■

Theorem 11.4 *Assume Condition (A) holds.*

a. *If there is a recurrent point, then the diffusion is recurrent.*
b. *If there is no recurrent point, then the diffusion is transient.*

Proof To prove part (a), suppose y is a recurrent point. Choose $0 < r_0 < r_1$, $z \in \mathbb{R}^k$ such that $|y - z| = r_1$. In Proposition 11.3(a⇒b), it is shown that

$$P_y(X(t) \in \partial B(z : r_0) \text{ for some } t \geq 0) = P_y(X(t) \in \overline{B(z : r_0)} \text{ for some } t \geq 0) = 1.$$

By Proposition 11.3(e), the diffusion is recurrent.

To prove part (b), suppose that no point in $\mathbb{R}^k$ is recurrent. Fix $x \in \mathbb{R}^k$, and let $r > |x|$ be an otherwise arbitrary positive number. By Proposition 11.3(e) and the maximum principle (Lemma 3(b)), for each $r_1 > r$ one has

$$\delta_{r_1} \equiv \sup_{|y|=r_1} P_y(\eta_{\overline{B(0:r)}^c} < \infty) < 1.$$

Define

$$\eta_1 = \inf\{t \geq 0 : X(t) \in \partial B(0 : r_1)\}, \eta_{2i} = \inf\{t \geq \eta_{2i-1} : X(t) \in \overline{B(0 : r)}\},$$

and

$$\eta_{2i+1} = \inf\{t \geq \eta_{2i} : X(t) \in \partial B(0 : r_1)\},\ i \geq 1.$$

By Lemma 7 and the strong Markov property, for all $i \geq 1$,

$$P_x(X(t) \in \overline{B(0 : r)} \text{ for some sequence of t's increasing to infinity})$$
$$\leq P_x(\eta_{2i+1} < \infty)$$

$$= \mathbb{E}_x(\mathbf{1}_{[\eta_{2i-1}<\infty]} P_{X(\eta_{2i-1})}(\eta_{\overline{B(0:r)}^c} < \infty))$$
$$\leq \delta_{r_1} P_x(\eta_{2i-1} < \infty). \tag{11.5}$$

Iterating this inequality, it follows that

$$P_x(X(t) \in \overline{B(0:r)} \text{ for some sequence of t's increasing to infinity}) \leq \delta_{r_1}^i, \forall i \geq 1.$$

Thus, $P_x(\liminf_{t\to\infty} |X(t)| > r) = 1$. Since $r > |x|$ is arbitrary, the proof is complete. ∎

We now turn to computable criteria to determine whether a multidimensional diffusion is recurrent or transient. However, this will require some additional notation.

Fix $r_0 > 0$ and for $x, z \in \mathbb{R}^k$, let $x' = x - z$.

(i)

$$A_z(x) = \sum_{i,j=1}^{k} a_{ij}(x'+z)\frac{x_i' x_j'}{|x'|^2}, \quad B_z(x) = 2\sum_{i=1}^{k} x_i' b_i(x'+z).$$

(ii)

$$C_z(x) = \sum_{i=1}^{k} a_{ii}(x'+z), \quad \overline{\beta}_z(r) = \sup_{|x'|=r} \frac{B_z(x) - A_z(x) + C_z(x)}{A_z(x)}.$$

(iii)

$$\underline{\beta}_z(r) = \inf_{|x'|=r} \frac{B_z(x) - A_z(x) + C_z(x)}{A_z(x)}, \quad \overline{\alpha}_z(r) = \sup_{|x'|=r} A_z(x).$$

(iv)

$$\underline{\alpha}_z(r) = \min_{|\lambda'|=r} A_z(x), \quad \overline{I}_z(r) = \int_{r_0}^{r} \frac{\overline{\beta}_z(u)}{u} du, \quad \underline{I}_z(r) = \int_{r_0}^{r} \frac{\underline{\beta}_z(u)}{u} du.$$

Lemma 8 *Let $F \in C^2((0,\infty))$, and let $z \in \mathbb{R}^k$. Define the function*

$$f(x) = F(|x-z|), \quad x \in \mathbb{R}^k, \ |x-z| > 0.$$

Then

$$2Lf(x) = A_z(x)F''(|x-z|) + \frac{F'(|x-z|)}{|x-z|}[B_z(x) - A_z(x) + C_z(x)].$$

Proof Straightforward differentiations yield for $|x-z|>0$,

$$\frac{\partial f(x)}{\partial x_i} = \frac{(x_i - z_i)}{|x-z|}F'(|x-z|)$$

$$\frac{\partial^2 f(x)}{\partial x_i^2} = \frac{(x_i - z_i)^2}{|x-z|^2}F''(|x-z|) - \frac{(x_i - z_i)^2}{|x-z|^3}F'(|x-z|) + \frac{F'(|x-z|)}{|x-z|}$$

$$\frac{\partial^2 f(x)}{\partial x_i x_j} = \frac{(x_i - z_i)(x_j - z_j)}{|x-z|^2}F''(|x-z|) - \frac{(x_i - z_i)(x_j - z_j)}{|x-z|^3}F'(|x-z|),\ i \neq j.$$

From here one may directly check the asserted expression for $2Lf(x)$. ■

Theorem 11.5 *Assume Condition (A) holds.*

a. *If for some $r_0 > 0$ and $z \in \mathbb{R}^k$, $\int_{r_0}^{\infty} \exp(-\overline{I}_z(r))dr = \infty$, then the diffusion associated with L is recurrent.*
b. *If for some $r_0 > 0$ and $z \in \mathbb{R}^k$, $\int_{r_0}^{\infty} \exp(-\underline{I}_z(r))dr < \infty$, then the diffusion associated with L is transient.*

Proof To prove part (a), assume that for some $r_0 > 0$ and $z \in \mathbb{R}^k$, $\int_{r_0}^{\infty} \exp(-\overline{I}_z(r))dr = \infty$. Define,

$$F(r) = \int_{r_0}^{r} \exp(-\overline{I}_z(u))du = \infty, \quad f(x) = F(|x-z|),\ |x-z| \geq r_0.$$

Let $z \in \mathbb{R}^k$ be such that $r = |x-z| > r_0$. Define stopping times by the escape times

$$\eta_{B(z:r_0)} = \inf\{t \geq 0 : X(t) \in \partial B(z:r_0)\}, \eta_N = \eta_{B(z:r_0)} \wedge \eta_{B(x:N)}. \tag{11.6}$$

By Theorem 11.1 and optional sampling,

$$M(t) = f(X(t \wedge \eta_N)) - \int_0^{t\wedge\eta_N} Lf(X(s))ds, \quad t \geq 0,$$

is a P_x-martingale, provided $|x-z| < N$. Thus, for $r = |x-z| > r_0$,

$$\begin{aligned}
&2\mathbb{E}_x F(|X(t\wedge\eta_N) - z|) - 2F(r) \\
&= \mathbb{E}_x \int_0^{t\wedge\eta_N} 2Lf(X(s))ds \\
&\leq \mathbb{E}_x \int_0^{t\wedge\eta_N} A_z(X(s))\{F''(|X(s)-z|) + \frac{F'(|X(s)-z|)}{|X(s)-z|}\overline{\beta}_z(|X(s)-z|)\}ds \\
&= 0,
\end{aligned} \tag{11.7}$$

since

$$F'(u) \geq 0, \quad F''(u) + \frac{1}{u} F'(u) \overline{\beta}_z(u) = 0, \quad u \geq r_0.$$

Letting $t \uparrow \infty$ in (11.6), and recalling that $\eta_N < \infty$ P_x-a.s., one obtains

$$\mathbb{E}_x F(|X(\eta_N) - z|) \leq F(r) = \int_{r_0}^{r} \exp\{-\overline{I}_z(u)\} du. \tag{11.8}$$

On evaluating the left side of (11.8), one has

$$P_x(\eta_{B(z:r_0)} > \eta_{B(x:N)}) \int_{r_0}^{N} \exp\{-\overline{I}_z(u)\} du \leq \int_{r_0}^{r} \exp\{-\overline{I}_z(u)\} du. \tag{11.9}$$

Now let $N \uparrow \infty$, to obtain

$$P_x(\eta_{B(z:r_0)} = \infty) \leq \lim_{N \uparrow \infty} \frac{\int_{r_0}^{r} \exp\{-\overline{I}_z(u)\} du}{\int_{r_0}^{N} \exp\{-\overline{I}_z(u)\} du} = 0. \tag{11.10}$$

Thus, $P_x(\eta_{B(z:r_0)} < \infty) = 1$, and the diffusion is recurrent by Proposition 11.3(e).

To prove part (b), assume that for some $r_0 > 0$ and $z \in \mathbb{R}^k$, $\int_{r_0}^{\infty} \exp(-\underline{I}_z(r)) dr < \infty$. Define

$$G(r) = \int_{r_0}^{r} \exp\{-\underline{I}_z(u)\} du, \quad g(x) = G(|x - z|), |x - z| \geq r_0.$$

Since $G'(u) \geq 0$ and $G''(u) + \frac{1}{u} G'(u) \underline{\beta}_z(u) = 0$ for $u \geq r_0$, one obtains, as above,

$$\mathbb{E}_x G(|X(\eta_N) - z|) - G(|x - z|) \geq 0,$$

or

$$P_x(\eta_{B(z:r_0)} > \eta_{B(x:N)}) \int_{r_0}^{N} \exp\{-\underline{I}_z(u)\} du \geq \int_{r_0}^{|x-z|} \exp\{-\underline{I}_z(u)\} du. \tag{11.11}$$

Hence, letting $N \uparrow \infty$, one has

$$P_x(\eta_{B(z:r_0)} = \infty) \geq \frac{\int_{r_0}^{|x-z|} \exp\{-\underline{I}_z(u)\} du}{\int_{r_0}^{\infty} \exp\{-\underline{I}_z(u)\} du} > 0.$$

Thus, the diffusion is not recurrent by Proposition 11.3, and therefore it is transient by Theorem 11.4. ■

We may now use a *renewal or regeneration method* to provide conditions for the existence of a unique invariant probability (see Bhattacharya and Waymire (2022),

Chapter 16, in the discrete parameter case). Recall (11.2) for the definition of the escape time η_U.

Lemma 9 *Assume Condition (A) holds.*

a. *The diffusion is recurrent and admits a finite invariant measure if there is a* $z \in \mathbb{R}^k$ *such that*

$$\sup_{y\in\partial B(z:r_1)} \mathbb{E}_y \eta_{\overline{B(z:r_0)}^c} < \infty,$$

for some $0 < r_0 < r_1$.

b. *Assume that the process is recurrent, but there is a* $z \in \mathbb{R}^k$ *and* $0 < r_0 < r_1$ *such that*

$$\mathbb{E}_y \eta_{\overline{B(z:r_0)}^c} = \infty,$$

for all $y \in \partial B(z : r_1)$. *Then there does not exist a finite invariant measure.*

Proof Consider part (a) first, and assume there is a $z \in \mathbb{R}^k$ such that

$$\sup_{y\in\partial B(z:r_1)} \mathbb{E}_y \eta_{\overline{B(z:r_0)}^c} < \infty,$$

for some $0 < r_0 < r_1$. Then the diffusion is recurrent by Proposition 11.3(e). Define

$$\eta_1 \equiv \eta_1(X) = \inf\{t \geq 0 : X(t) \in \partial B(z : r_0)\},$$

$$\eta_{2i} \equiv \eta_{2i}(X) = \inf\{t \geq \eta_{2i-1} : X(t) \in \partial B(z : r_1)\},$$

and

$$\eta_{2i+1} \equiv \eta_{2i+1}(X) = \inf\{t \geq \eta_{2i} : X(t) \in \partial B(z : r_0)\},\ i \geq 1.$$

Under the assumption of part (a) and the strong Markov property, the sequence $\tilde{X}_0, \tilde{X}_i = X(\eta_{2i+1}), i = 0, 1, \ldots$ is a recurrent Markov chain on the compact state space $S = \partial B(z : r_0)$, and therefore has an invariant probability $\tilde{\pi}(dx)$ for the Feller transition probabilities[5] given by

$$\tilde{p}(x, A) = \int_{\partial B(z:r_1)} P_y(X(\eta_3) \in A) P_x(X(\eta_2) \in dy),$$

for $A \in \mathcal{B}(S), x \in S = \partial B(z : r_0)$. In particular, $\tilde{\pi}$ extends to Borel measurable sets $C \subset \mathbb{R}^k$, by defining a sigma-finite measure by

[5] See Bhattacharya and Waymire (2022), Proposition 8.6.

$$\pi(C) = \int_S \mathbb{E}_x \rho_C \tilde{\pi}(dx), \tag{11.12}$$

where

$$\rho_C = \int_{\eta_1}^{\eta_3} \mathbf{1}_C(X(t))dt$$

denotes the occupation time of C by the process X during a single cycle. Equivalently, letting ρ denote the time to complete a cycle starting in $S = \partial B(z : r_0)$, π is defined by the condition that for any bounded continuous function f on $\mathbb{R}^k$,

$$\int_{\mathbb{R}^k} f(x)\pi(dx) = \int_S \mathbb{E}_x \int_0^{\rho} f(X(s))ds\tilde{\pi}(dx). \tag{11.13}$$

To prove that π is an invariant probability we use the following formula of Khaśminskii (1960). For a random variable $g(X)$, and a stopping time τ such that $\mathbb{E}_x \int_0^\tau g(X_t^+)dt$ exists, where X_t^+ denotes the after-t process defined by $X_t^+(s) = X(t+s)$, $s \geq 0$,

$$\mathbb{E}_x \int_0^\tau \mathbb{E}_{X(t)} g(X)dt = \mathbb{E}_x \int_0^\tau g(X_t^+)dt. \tag{11.14}$$

To check (11.14), apply Fubini's theorem to the right side and then condition on $\mathcal{F}_t$ to get,

$$\begin{aligned}
\mathbb{E}_x \int_0^\tau g(X_t^+)dt &= \mathbb{E}_x \int_0^\infty \mathbf{1}_{[\tau>t]} g(X_t^+)dt \\
&= \int_0^\infty \mathbb{E}_x \mathbf{1}_{[\tau>t]} g(X_t^+)dt \\
&= \int_0^\infty \mathbb{E}_x \left\{ \mathbb{E}\big(\mathbf{1}_{[\tau>t]} g(X_t^+)|\mathcal{F}_t\big) \right\} dt \\
&= \int_0^\infty \mathbb{E}_x \mathbf{1}_{[\tau>t]} \{\mathbb{E}\big(g(X_t^+)|\mathcal{F}_t\big)\} dt \\
&= \mathbb{E}_x \int_0^\tau \mathbb{E}_{X(t)} g(X)dt.
\end{aligned} \tag{11.15}$$

Now, with $\tau = \rho$, $h \in C_0(\mathbb{R}^k)$, $f(x) = \mathbb{E}_x h(X(t))$ in (11.13), and using (11.14) as well,

$$\int_{\mathbb{R}^k} T_t h(x)\pi(dx) = \int_{\mathbb{R}^k} \mathbb{E}_x h(X(t))\pi(dx)$$

$$
\begin{aligned}
&= \int_S \mathbb{E}_x \int_0^\rho \mathbb{E}_{X(s)} h(X(t)) ds \tilde{\pi}(dx) \\
&= \int_S \mathbb{E}_x \int_0^\rho h(X_s^+(t)) ds \tilde{\pi}(dx) \\
&= \int_S \mathbb{E}_x \int_0^\rho h(X(t+s)) ds \tilde{\pi}(dx) \\
&= \int_S \mathbb{E}_x \int_t^{\rho+t} h(X(u)) du \tilde{\pi}(dx) \\
&= \int_S \mathbb{E}_x \int_0^\rho h(X(u)) du \tilde{\pi}(dx) \\
&+ \int_S \mathbb{E}_x \int_\rho^{\rho+t} h(X(u)) du \tilde{\pi}(dx) - \int_S \mathbb{E}_x \int_0^t h(X(u)) du \tilde{\pi}(dx)
\end{aligned}
\tag{11.16}
$$

Using (11.13), one has

$$
\begin{aligned}
\int_S \mathbb{E}_x \int_\rho^{\rho+t} h(X(u)) du \tilde{\pi}(dx) &= \int_S \mathbb{E}_x \mathbb{E}_{\tilde{X}_1} \int_0^t h(X(u)) du \tilde{\pi}(dx) \\
&= \int_S \mathbb{E}_x \int_0^t h(X(u)) du \tilde{\pi}(dx).
\end{aligned}
\tag{11.17}
$$

Thus, by (11.16) and (11.17), and again using (11.13),

$$
\int_{\mathbb{R}^k} T_t h(x) \pi(dx) = \int_S \mathbb{E}_x \int_0^\rho h(X(u)) du \tilde{\pi}(dx) = \int_{\mathbb{R}^k} h(x) \pi(dx). \tag{11.18}
$$

This proves the invariance of π as defined by (11.12).

To prove part (b) observe that if the expected value (of the return time) is infinite, starting from the boundary of the large sphere, then the invariant measure constructed in part (a) is not finite. Since every sigma-finite invariant measure is a multiple of this one, as follows from Proposition 11.6 or Theorem 11.7, the diffusion has no finite invariant measure. ■

For further analysis, especially concerning the convergence of the positive recurrent diffusion to equilibrium, we consider the discrete parameter Markov chain $\{X(n) : n = 0, 1, \ldots\}$, called the *discrete skeleton* of the diffusion $\{X(t) : t \geq 0\}$. It is defined to be recurrent or transient as in Definition 11.3, but replacing t by n. We have the following corollary to Proposition 11.3 and Theorem 11.4.

Proposition 11.6 *Assume Condition (A). If the diffusion is recurrent, then so is its discrete skeleton* $\{X(n) : n \geq 0\}$*, and if the diffusion is transient then so is* $\{X(n) : n \geq 0\}$.

Proof Let x_0 be a transient point of the diffusion. Then, by definition, so is it for the discrete skeleton. If x_0 is not a transient point of the diffusion, then there exists $N > 0$ such that $x_0 \in B(0 : N)$ and that $P_y(\overline{B(0 : N)}$ is reached in finite time$) = 1$, for all $y \in \mathbb{R}^k$. For arbitrary fixed $\varepsilon > 0$, let D be the event that $X(n)$ reaches $\overline{B(0 : \delta)}$, where $0 < \delta < \varepsilon$, at some time $n \geq 1$. By the strong Feller property, $y \to P_y(D)$ is continuous and positive on $\overline{B(0 : N)}$ whose minimum is denoted by $\lambda > 0$. If $\lambda < 1$, then conditioning on the event D^c and using the strong Markov property, the conditional probability that $\overline{B(0 : \delta)}$ is reached, after reaching $\overline{B(0 : N)}$ is again at least λ. Continuing in this way, the probability that by the nth try $\overline{B(0 : \delta)}$ is reached is $\sum_{1 \leq m \leq n-1} \lambda(1 - \lambda)^{m-1}$, which converges to 1 as $n \to \infty$. This shows that $P_{x_0}(X(n), n \geq 1$, reaches $B(x_0 : \varepsilon)) = 1$. By the strong Markov property, it follows that x_0 is a recurrent point for $\{X(n) : n \geq 0\}$. The argument actually shows that $\lambda = 1$. ■

Remark 11.4 Note that $P_y(D) = \mathbb{E}_y f(X(1))$, where

$$f(x) = P_x(X(n) \text{ reaches } \overline{B(0 : \delta)} \text{ for some } n \geq 0).$$

Therefore, $y \to P_y(D)$ is continuous by the strong Feller property.

Focusing on the discrete skeleton is especially useful for the following important result.

Theorem 11.7 *Assume Condition (A). (a) Suppose $\{X(t) : t \geq 0\}$, is a recurrent diffusion, and that there exists a measure ν such that for some $x_0 \in \mathbb{R}^k$ and $r > 0$, $\nu(\overline{B(x_0 : r)}) > 0$ and*

$$p(1; x, A) \geq \nu(A) \quad \forall\, x \in \overline{B(x_0 : r)},\ A \in \mathcal{B}(\mathbb{R}^k). \tag{11.19}$$

Then the diffusion has a sigma-finite invariant measure π, unique up to a scalar multiple. (b) In addition to the assumption in (a), assume the diffusion satisfies

$$\sup_{|x - x_0| = r_1} \mathbb{E}_x \tau_1 < \infty, \tag{11.20}$$

for any $x_0 \in \mathbb{R}^k$ and $r_1 > r > 0$, where $\tau_1 = \inf\{t \geq 0 : X(t) \in \overline{B(x_0 : r)}\}$ is the escape time from $B(x_0 : r)$. Then the diffusion has a unique invariant probability π and, moreover, whatever be the initial distribution, the distribution of the after-t process $X_t^+(s) = \{X(t + s) : s \geq 0\}$ converges in total variation norm to P_π, as $t \to \infty$.

Proof (a)We have proved that if the diffusion is recurrent so is $\{X(n) : n \geq 0\}$. The condition (11.19) implies that $\{X(n) : n \geq 0\}$, is Harris recurrent, and by Theorem 20.4 in Bhattacharya and Waymire (2022), it has a sigma-finite invariant measure π, unique up to a scalar multiple. Since every such measure is also invariant for the diffusion $\{X(t) : t \geq 0\}$ (Exercise 12), part (a) is proved. To prove part (b), we again consider the discrete process $\{X(n) : n \geq 0\}$, and prove the corresponding result

for it. Note that, given $\sup_{|x-x_0|=r_1} \mathbb{E}_x\tau_1 < \infty$ for the diffusion, one needs to check $\sup_{|x-x_0|=r_1} \mathbb{E}_x\eta < \infty$ where $\eta = \inf\{n > 0 : X(n) \in \overline{B(x_0 : r)}\}$, for given x_0, $0 < r < r_1$. Consider the following stopping times: $\tau_0 = 0$, $\tau_1 := \inf\{t > 0 : X(t) \in \overline{B(x_0, r)}\}$, $\tau_2 = \inf\{t > \tau_i : X(t) \in \partial B(x_0 : r_1)\}$, $\tau_{2i+1} = \inf\{t > \tau_{2i} : X(t) \in \overline{B(x_0 : r)}\}$, $\tau_{2i+2} = \inf\{t > \tau_{2i+1} : X(t) \in \partial B(x_0 : r_1)\}$, $(i = 0, 1, \dots)$. Choose r', $0 < r < r' < r_1$. Define $\tau'_{2i+1} = \inf\{t > \tau_{2i+1} : X(t) \in \overline{B(x_0 : r')}$. Let $E_i = \{\tau'_{2i+1} - \tau_{2i+1} \geq 1\}$, $i \geq 0$. Clearly, $\eta \leq \tau_{2i+2}$ on E_i; i.e., $\eta \leq \tau_{2\theta+2}$, where $\theta := \inf\{i \geq 0 : \tau'_{2i+1} - \tau_{2i+1} \geq 1\}$. By the strong Feller property, $\delta := \min\{P_x(E_0) : |x - x_0| = r_1\} > 0$. Hence $\mathbb{E}_x\eta \leq \mathbb{E}_x\tau_1 + \sum_{n=0}^{\infty}(1-\delta)^n \sup_{|x-x_0|=r} \mathbb{E}_x\tau_{2n+2} \leq \mathbb{E}_x\tau_1 + \sum_{n=0}^{\infty}(1-\delta)^n n \sup_{|x-x_0|=r_1} \mathbb{E}_x(\tau_2 - \tau_0)$. Note that $C = \sup_{|x-x_0|=r_1} \mathbb{E}_x\tau_1 < \infty$ by hypothesis. This implies $\sup_{|x-x_0|=r_1} \mathbb{E}_x(\tau_{2i+1} - \tau_{2i}) \leq C, i \geq 1$. Also, $\sup_{|x-x_0|=r_1} \mathbb{E}_x(\tau_{2i+2} - \tau_{2i+1})$ is finite, $i \geq 1$, since the escape time from a compact set has finite moments of all orders, uniformly for all initial x in the set. Thus we have proved $\sup_{x \in \overline{B(x_0:r')}} \mathbb{E}_x\eta < \infty$. By Theorem 20.4 in Bhattacharya and Waymire (2022), since $\{X(n) : n \geq 0\}$ is locally minorized on $\mathcal{A}_0 = \overline{B(x_0 : r')}$, the discrete time process $\{X(n) : n \geq 0\}$, has a unique invariant probability π and whatever be the initial distribution μ, the distribution Q_n of the after n-process $X_n^+ = \{X(n+m) : m = 0, 1, \dots\}$ converges in total variation distance to the distribution $\tilde{P}_\pi$, say, of the stationary distribution of $\{X(n) : n \geq 0\}$, under the initial distribution π. The desired result (b) for the diffusion easily follows, because the diffusion has a unique invariant probability and the convergence of X_t^+ in total variation as $t \to \infty$ is an easy consequence of the convergence of X_n^+ as $n \to \infty$. ■

Remark 11.5 From standard results in PDE (See, e.g., Friedman (1964)), the local minorization (11.19) holds in the case the coefficients of L are Hölder continuous (Exercise 9), ensuring Harris recurrence of $\{X(n) : n \geq 0\}$. We expect something similar to be true under Condition (A) and recurrence to imply Harris recurrence using the estimates for the transition probability density in Stroock and Varadhan (1979), Chapter 9, Section 2.

We now turn to verifiable conditions for null and positive recurrence based on the coefficients of L.

Theorem 11.8 *Assume Condition (A) holds.*

a. *If for some $r_0 > 0$ and $z \in \mathbb{R}^k$,*

$$\int_{r_0}^{\infty} \exp(-\overline{I}_z(r))dr = \infty, \quad \int_{r_0}^{\infty} \frac{1}{\underline{\alpha}_z(r)} \exp(-\overline{I}_z(r))dr < \infty,$$

then there exists a finite invariant measure (unique up to constant multiples) for the diffusion associated with L.

b. *If for some $r_0 > 0$ and $z \in \mathbb{R}^k$,*

$$\int_{r_0}^{\infty} \exp(-\overline{I}_z(r))dr = \infty, \quad \lim_{N\to\infty} \frac{\int_{r_0}^{N} \exp(-\overline{I}_z(s))(\int_{r_0}^{s} (\overline{I}_z(r))/\overline{\alpha}_z(r))dr)ds}{\int_{r_0}^{N} \exp(-\underline{I}_z(r))dr} = \infty,$$

then the recurrent diffusion does not admit a finite invariant measure.

Proof To prove (a), assume for some $r_0 > 0$ and $z \in \mathbb{R}^k$, $\int_{r_0}^{\infty} \exp(-\overline{I}_z(r))dr = \infty$ and $\int_{r_0}^{\infty} \frac{1}{\underline{\alpha}_z(r)} \exp(-\overline{I}_z(r))dr < \infty$. Define,

$$F(r) = -\int_{r_0}^{r} \exp(-\overline{I}_z(s))(\int_{s}^{\infty} \frac{1}{\underline{\alpha}_z(u)} \exp\{\overline{I}_z(u)\}du)ds, \quad r > r_0.$$

Then, for $r \geq r_0$,

$$F'(r) = -\exp\{-\overline{I}_z(r)\} \int_{r}^{\infty} \frac{1}{\underline{\alpha}_z(u)} \exp\{\overline{I}_z(u)\}du < 0,$$

and

$$F''(r) = -\frac{\overline{\beta}_z(r)}{r} F'(r) + \frac{1}{\underline{\alpha}_z(r)}.$$

Let

$$f(x) = F(|x - z|), \quad |x - z| \geq r_0.$$

Then, using these last two calculations,

$$2LF(x) \geq \frac{A_z(x)}{\underline{\alpha}_z} \geq 1, \quad |x - z| \geq r_0.$$

If $\eta_{B(x_0;r)}$, η_N are as in (11.6), then as in the proof of Theorem 11.5, one has

$$\begin{aligned} 2\mathbb{E}_x F(|X(t \wedge \eta_N - z|) - 2F(|x - z|) &= \mathbb{E}_x \int_0^{t\wedge\eta_N} 2Lf(X(s))ds \\ &\geq \mathbb{E}_x(t \wedge \eta_N), \quad r_0 \leq |x - z| \leq N. \end{aligned}$$

First, letting $t \uparrow \infty$ and then $N \uparrow \infty$, from here one has, since $\eta_N \uparrow \eta_{\overline{B(x:\eta_0)}^c}$ P_x-a.s. (due to recurrence),

$$\mathbb{E}_x \eta_{\overline{B(x:\eta_0)}^c} \leq -2F(|x - z|).$$

Now apply Lemma 9(a).

To prove (b) assume that for some $r_0 > 0$ and $z \in \mathbb{R}^k$, $\int_{r_0}^{\infty} \exp(-\overline{I}_z(r))dr = \infty$, and $\lim_{N\to\infty} \frac{\int_{r_0}^{N} \exp(-\overline{I}_z(s))(\int_{r_0}^{s}(\overline{I}_z(r))/\overline{\alpha}_z(r))dr)ds}{\int_{r_0}^{N} \exp(-\underline{I}_z(r)dr} = \infty$. Define,

$$G(r) = \int_{r_0}^{r} \exp(-\overline{I}_z(s))(\int_{r_0}^{s} \frac{1}{\overline{\alpha}_z(u)} \exp\{\overline{I}_z(u)\}du)ds, \quad |x-z| > r_0,$$

and

$$g(x) = G(|x-z|) \quad |x-z| > r_0.$$

Then, for $r \geq r_0$, $G'(r) \geq 0$, and $G''(r) = -\frac{\overline{\beta}_z(r)}{r}G'(r) + \frac{1}{\overline{\alpha}_z(r)}$, so that $2Lg(x) \leq 1$. Therefore one has

$$\mathbb{E}_x(t \wedge \eta_N) \geq 2\mathbb{E}_x G(|X(t \wedge \eta_N) - z|) - 2G(|x-z|). \tag{11.21}$$

Thus, letting $t \uparrow \infty$, from here one obtains

$$\begin{aligned}\mathbb{E}_x \eta_N &\geq 2\mathbb{E}_x G(|X(\eta_N) - z|) - 2G(|x-z|)\\ &= 2P_x(\eta_{B(x:N)} < \eta_{\overline{B(x:\eta_0)}^c})G(N) - 2G(|x-z|).\end{aligned}$$

Now let $N \uparrow \infty$, using (11.11),

$$\mathbb{E}_x \eta \geq 2 \lim_{N\to\infty} \frac{\int_{r_0}^{|x-z|} \exp\{-\underline{I}_z(u)\}du}{\int_{r_0}^{N} \exp\{-\underline{I}_z(u)\}du} G(N) - 2G(|x-z|) = \infty.$$

The proof is now complete by Lemma 9(b). ■

Remark 11.6 The null recurrence criterion[6] presented in Theorem 11.8(b) does not quite match with the one announced by Khaśminskii, which we are unable to verify.

In view of the criteria for transience and recurrence provided by Theorem 8.8, we will focus on criteria for null and positive recurrence to complete the picture in one dimension.

Definition 11.6 Let

$$\tau_y^x := \inf\{t \geq 0 : X(t) = y\},\ X(0) = x, \quad x, y \in \mathbb{R}.$$

A recurrent diffusion is said to be *positive recurrent* (or *ergodic*) if $\mathbb{E}\tau_a^x < \infty\ \forall\ x, a$, and *null recurrent* if $\mathbb{E}\tau_a^x = \infty$ for some x, a.

[6] To date the authors are unaware of a proof nor counterexample for Khaśminskii's announced condition for null recurrence.

One-dimensional diffusions admit a very special representation in terms of the direction and rate at which they exit subintervals of $\mathbb{R}$. As is more fully exploited in Chapter 21, one of Feller's great innovations for one-dimensional diffusion was to recognize the essential roles of scale and speed functions defined as follows. Fix an arbitrary $x_0 \in \mathbb{R}$ (e.g., $x_0 = 0$), and define the *scale function*

$$s(x) \equiv s(x_0; x) := \int_{x_0}^{x} \exp\{-I(x_0, z)\} dz, \tag{11.22}$$

where

$$I(x, z) := \int_{x}^{z} \{2\mu(y)/\sigma^2(y)\} dy,$$

with the convention

$$I(z, x) = -I(x, z), \quad \text{and} \quad I(x) \equiv I(0, x).$$

Also define the *speed function*

$$m(x) \equiv m(x_0; x) := \int_{x_0}^{x} (2/\sigma^2(z)) \exp\{I(x_0, z)\} dz. \tag{11.23}$$

With this now observe that one may express the form of the differential operator

$$A = \frac{1}{2}\sigma^2(x)\frac{d^2}{dx^2} + \mu(x)\frac{d}{dx}, \tag{11.24}$$

as

$$A = \frac{d}{dm(x)} \frac{d}{ds(x)}. \tag{11.25}$$

Here one defines $\frac{df(x)}{ds(x)} = \frac{df(x)}{dx} \Big/ \frac{ds(x)}{dx} = f'(x)/s'(x)$, and $\frac{df(x)}{dm(x)} = f'(x)/m'(x)$ according to the familiar chain rule from calculus.

Next observe that the solution of the *two-point boundary value problem* (8.47):

$$A\varphi(x) = 0 \text{ for } c < x < d, \quad \varphi(c) = 0, \ \varphi(d) = 1,$$

is simply $c_1 s(x) + c_2$. In particular, one may apply the boundary conditions to obtain the probability of escaping to the right given in (8.49) in the equivalent form

$$P(\tau_d^x < \tau_c^x) = \frac{s(d) - s(x)}{s(d) - s(c)}, \quad c \le x \le d. \tag{11.26}$$

Similarly

$$P(\tau_c^x < \tau_d^x) = \frac{s(x) - s(c)}{s(d) - s(c)}, \quad c \le x \le d. \tag{11.27}$$

It also now follows that the conditions (8.53) for recurrence may be expressed as $s(-\infty) = -\infty$ and $s(\infty) = \infty$.

Next consider another two-point boundary value problem.

$$\begin{aligned} &A\gamma(x) = -1 \\ &\gamma(c) = 0, \ \gamma(d) = 0, \end{aligned} \qquad c < x < d, \tag{11.28}$$

where A is the operator $\frac{1}{2}\sigma^2(x)\frac{d^2}{dx^2} + \mu(x)\frac{d}{dx}$. By Itô's lemma,

$$Z(t) := \gamma(X^x(t)) - \int_0^t -1 ds \equiv \gamma(X^x(t)) + t \qquad (t \ge 0), \tag{11.29}$$

is a martingale. With $\tau_\partial^x \equiv \tau_{\partial(c,d)}^x := \tau_c^x \wedge \tau_d^x$ $(c \le x \le d)$, $\{Z(\tau_\partial^x \wedge t) : t \ge 0\}$ is bounded by $\max\{|\gamma(y)| : c \le y \le d\} + \tau$, and is therefore uniformly integrable, since by Lemma 8.4 $\mathbb{E}\tau_\partial^x < \infty$. Applying the optional stopping theorem one obtains, noting that $\gamma(X^x(\tau_\partial^x)) = 0$,

$$\mathbb{E}\tau_\partial^x = \gamma(x). \tag{11.30}$$

Observe that one may express the solution of the *two-point boundary value problem* (11.28) as $\gamma(x) = c_1 s(x) - \int_c^x m(y)ds(y) + c_2$. Applying the boundary conditions $\gamma(c) = \gamma(d) = 0$, yields the solution in the form

$$\mathbb{E}\tau_\partial^x = \gamma(x) = \frac{s(c;x)}{s(c;d)} \int_c^d m(c;y)s'(c;y)dy - \int_c^x m(c;y)s'(c;y)dy. \tag{11.31}$$

Thus the criteria for positive recurrence may be expressed as that of recurrence $s(-\infty) = -\infty$, $s(\infty) = \infty$, together with

$$\mathbb{E}\tau_c^x = \int_c^x (m(\infty) - m(y))ds(y) < \infty \ \forall\, x > c \iff m(\infty) < \infty, \tag{11.32}$$

and

$$\mathbb{E}\tau_d^x = \int_x^d (m(y) - m(-\infty))ds(y) < \infty \ \forall\, x < d. \iff m(-\infty) > -\infty. \tag{11.33}$$

The next result is a one-dimensional version of Theorem 11.7.

Theorem 11.9 *Suppose $\mu(\cdot), \sigma(\cdot)$, are locally Lipschitz and $\sigma^2(x) > 0$ for all x. (a) Let $s(0) = 0$. If $s(\infty) = \infty$ and $s(-\infty) = -\infty$, then the diffusion is recurrent; if one of $s(-\infty)$, $s(\infty)$ is finite, then the diffusion is transient. (b) If the diffusion is recurrent, but at least one of the integrals in (11.32), (11.33) diverges, i.e., if at least one of $m(-\infty)$, $m(\infty)$ is not finite, then the diffusion is null recurrent and it has a sigmafinite invariant measure which is unique up to scalar multiples. (c) Suppose the diffusion is recurrent. If both integrals in (11.32) (11.33) converge, i.e., if $m(-\infty)$ and $m(\infty)$ are both finite, then there is a unique invariant probability, and the after-t process X_t^+ converges in total variation distance to P_π as $t \to \infty$, no matter what the initial distribution may be.*

Proof We have proved the criteria for recurrence, transience, and of null and positive recurrence above. In order to prove the uniqueness and convergence assertions in (b) and (c), apply the one-dimensional analog of Theorem 11.7 as follows. Let $a < b < c < d$. If the diffusion is recurrent, then $[a, b]$ is a recurrent set. Also, the local minorization condition (11.19) holds, i.e., $p(1; x, A) \geq \nu(A)$ for all $x \in [a, b]$, A a Borel subset of $[a, b]$, and $\nu(A) = \int_A p(1; x_0, y)dy$, where $p(1; x_0, y) = \min_{x\in[a,b]} p(1; x, y) > 0$ (see Exercise 9). Hence part (b)follows from Theorem 11.7, part (a). For part (c), define the stopping times $\tau_1 = \inf\{t \geq 0 : X(t) \in [c, d]\}$, $\tau_2 = \inf\{t > \tau_1, X(t) \in [a, b]\}$, $\tau_{2i+1} = \inf\{t > \tau_{2i}, X(t) \in [c, d]\}$, $\tau_{2i+2} = \inf\{t > \tau_{2i+1} : X(t) \in [a, b]\}$ $(i = 1, 2, \ldots)$, $\tau_z^x = \inf\{t > 0 : X(t) = z, X(0) = x\}$. Then $\sup_{y\in[c,d]} \mathbb{E}_y\tau_2 = \mathbb{E}\tau_b^d < \infty$, by (11.32). Similarly, $\sup_{y\in[a,b]} \mathbb{E}_y\tau_1 = \mathbb{E}\tau_c^a < \infty$, by (11.33). Going through the same argument over cycles $[\tau_{2i}, \tau_{2i+2}] (i \geq 1)$, as in the proof of part (b) of Theorem 11.7, the proof is complete. ■

Proposition 11.10 *Suppose $\mu(\cdot), \sigma(\cdot)$ are locally Lipschitzian with $\sigma^2(x) > 0$ for all x. Then the diffusion $X = \{X(t) : t \geq 0\}$ admits a unique invariant probability if it is positive recurrent, and the invariant probability is absolutely continuous with respect to Lebsgue measure, i.e. $\pi(dx) = \pi(x)dx$, where the density $\pi(x)$ is given by*

$$\pi(x) = \theta^{-1} \cdot \frac{1}{\sigma^2(x)} e^{I(0,x)} = \theta^{-1} m'(x), \qquad -\infty < x < \infty, \tag{11.34}$$

where θ^{-1} is the normalizing constant, and m is the speed function defined by (11.23) with $x_0 = 0$.

Proof To prove absolute continuity of the invariant probability $\pi(dx)$, note that if $\lambda(B) = 0$, where λ is Lebesgue measure on $\mathbb{R}$, then $p(1; x, B) = 0$ for all $x \in \mathbb{R}$. Hence $\pi(B) = \int_{\mathbb{R}} p(1; x, B)\pi(dx) = 0$. Next, fix $c < d$. For any initial $X(0)$, define

$$\eta_1 \equiv \eta_1(X) := \tau_c(X) = \inf\{t \geq 0 : X(t) = c\}, \ \eta_2 := \inf\{t > \eta_1 : X(t) = d\},$$
$$\tag{11.35}$$
$$\eta_{2n+1} := \inf\{t > \eta_{2n} : X(t) = c\}, \quad \eta_{2n+2} := \inf\{t > \eta_{2n+1} : X(t) = d\}, \ (n \geq 1).$$

Since $(\eta_{2i+1} - \eta_{2i-1})$ $(i = 1, 2, \dots)$ are i.i.d., it easily follows from the strong law of large numbers that the unique invariant probability is given by

$$\pi(B) := \frac{1}{\mathbb{E}(\eta_3 - \eta_1)} \mathbb{E} \int_{\eta_1}^{\eta_3} \mathbf{1}_B(X(s))ds \qquad B \in \mathcal{B}(\mathbb{R}), \tag{11.36}$$

To show that $\pi(x)$ in (11.34) is the invariant density, let f be a twice continuously differentiable nonnegative function on $\mathbb{R}$ with compact support. By (11.36), letting $\pi(dy)$ be the invariant probability,

$$\begin{aligned}
\int_{\mathbb{R}} f(y)\pi(dy) &= \frac{1}{\mathbb{E}(\eta_3 - \eta_1)} \mathbb{E} \int_{\eta_1}^{\eta_3} f(X(s))ds \\
&= c_0 \mathbb{E}[\int_{\eta_1}^{\eta_2} f(X(s))ds + \int_{\eta_2}^{\eta_3} f(X(s))ds] \\
&= c_0[\mathbb{E} \int_0^{\tau_d} f(X^c(s))ds + \mathbb{E} \int_0^{\tau_c} f(X^d(s))ds] \\
&= c_0[\varphi_1(d) + \varphi_2(c)],
\end{aligned} \tag{11.37}$$

where $c_0 := \frac{1}{\mathbb{E}(\eta_3 - \eta_1)}$, say, and, writing X^x to denote the process X starting at x, by the optional stopping rule applied to the martingales $\varphi_i(X^x(t)) - \int_0^t A\varphi_i(X^x(s))ds$, $i = 1, 2$, $\varphi_1(x) := \mathbb{E} \int_0^{\tau_c} f(X^x(s))ds$ $(x \geq c)$, and $\varphi_2(x) := \mathbb{E} \int_0^{\tau_d} f(X^x(s))ds$ $(x \leq d)$, where φ_1, φ_2 are the solutions of

$$\begin{aligned}
&A\varphi_1(x) = -f(x) \ (x > c),\ \varphi_1(c) = 0,\ \ \varphi_1(x) \to \text{ finite limit as } x \uparrow +\infty, \\
&A\varphi_2(x) = -f(x) \ (x < d), \varphi_2(d) = 0,\ \ \varphi_2(x) \to \text{ finite limit as } x \downarrow -\infty.
\end{aligned} \tag{11.38}$$

Now suppose that f vanishes off (a, b) and, use $A = \frac{d}{dm}\frac{d}{ds}$ and the boundary conditions at c, d, respectively, to integrate (11.38) as

$$\varphi_1(x) = \int_c^x \{\int_z^b f(y)m'(y)dy\}s'(z)dz \tag{11.39}$$

and

$$\varphi_2(x) = \int_x^d \{\int_a^z f(y)m'(y)dy\}s'(z)dz \tag{11.40}$$

Now simply check that

$$c_0[\varphi_1(d) + \varphi_2(c)] = \int_{\mathbb{R}} f(x)\pi(x)dx, \quad \pi(x) := \theta \frac{1}{\sigma^2(x)} e^{I(0,x)}. \tag{11.41}$$

The result now follows from (11.37). ■

Remark 11.7 Since $\frac{d}{dt}\int_{\mathbb{R}} T_t f(x)\pi(dx) = 0$ for all (smooth) bounded f, one way to look for the invariant density is to try to solve $\int_{\mathbb{R}} Af(x)\pi(x)dx = 0$. In particular, after an integration by parts (twice), one has $\int_{\mathbb{R}} f(y)A^*\pi(y)dy = 0$, where $A^*\pi(y) := \frac{\partial^2(\frac{1}{2}\sigma^2(y)\pi(y))}{\partial y^2} - \frac{\partial(\mu(y)\pi(y))}{\partial y}$. From here one may compute the asserted invariant probability in Proposition 11.10 by solving $A^*\pi = 0$ with decay at infinities (required for integrability of π).

Example 1 (Stochastic Logistic Model) The logistic model with additive noise is given by the Itô equation

$$dX(t) = rX(t)(k - X(t))dt + \sigma X(t)dB(t), \ X(0) = x > 0, (r > 0; k > 0; \sigma \neq 0) \tag{11.42}$$

If $\sigma = 0$ then it reduces to the deterministic logistic growth model[7] The state "0" is an absorbing state, i.e., if the process starts at 0 or reaches 0; then it stays there forever (it is the "state of extinction"). If $\sigma \neq 0$, the diffusion[8] does not reach zero, starting from $x > 0$. Hence we take the state space to be $(0, \infty)$. The value k is often called the *carrying capacity* of the population: it represents the population size that the environment (available resources) can carry or sustain. A pathwise solution to the SDE is given in Chapter 8, Example 7. The reader is invited to provide the proofs of the following results using the results of this chapter (Exercise 11).

A. $2rk/\sigma^2 < 1$: The process approaches 0 as $t \to \infty$, i.e., the population is "threatened with extinction" or "becomes endangered". The diffusion is transient, and the probability of actually hitting 0 in finite time is zero, if the process starts at $x > 0$.
B. $2rk/\sigma^2 = 1$: In this case the diffusion on $(0, \infty)$ is null recurrent.
C. $2rk/\sigma^2 > 1$: The diffusion $\{X(t) : t \geq 0\}$ is positive recurrent on $(0, \infty)$, and approaches a unique steady state as $t \to \infty$, no matter what the initial state may be.

With harvesting, the model is, for $r > 0$; $k > 0$, $t > 0$, $\theta > 0$, given by

$$dX(t) = rX(t)(k - X(t))dt + \theta dt + \sigma X(t)dB(t), \ X(0) = x > 0, \tag{11.43}$$

where the positive constant θ denotes the rate of consumption or harvesting. In this model $X(t) \to 0$ as $t \to \infty$, but never reaches zero. Next consider a model in which the harvesting rate is proportional to the stock, i.e., for $r > 0$, $k > 0$, $\sigma \neq 0$,

$$dX(t) = rX(t)(k - X(t))dt\theta X(t)dt + \sigma X(t)dB(t), \ X(0) = x > 0. \tag{11.44}$$

[7] See Bacaer (2011) for an interesting historical account of this model.

[8] The results for the stochastic model ($\sigma \neq 0$) are given in Bhattacharya and Majumdar (2022). Multiplicative random perturbations on population growth are treated in the Special Topics Chapter 26.

Then

a. The diffusion is transient if and only if $kr - \theta < \sigma^2/2$.
b. If $kr - \theta = \sigma^2/2$ the process is null recurrent, and
c. If $kr-\theta > \sigma^2/2$ the difusion $\{X(t) : t \geq 0\}$ is positive recurrent and it converges to a unique steady state (invariant probability).

For additional examples of compelling mathematical interest see those at the end of the special topics Chapter 14.

Remark 11.8 In the one-dimensional case, the asymptotic criteria for transience, recurrence, and positive recurrence in Theorem 11.9 hold with Feller's scale function $s(\cdot)$ and speed function $m(\cdot)$ (see Chapter 21, Definition 21.2) in place of the corresponding quantities defined by (11.22), (11.23), thus relaxing condition (A). In the multidimensional case, condition (A) may be relaxed a little by allowing isolated singularities, as long as the integral conditions in Theorems 11.5, 11.8 hold.

Exercises

1. Under the hypothesis of Lemma 3, prove that the distribution $p(x_0, dy)$ of $X(\tau_B)$ on ∂B has full support ∂B. [*Hint:* If $p(x_0, dy)$ does not have full support on ∂B, then there exists a non-empty open subset $O \subset \partial B$ such that $p(x_0, O) = 0$. Pick a point $z \in O$, and let $\omega_0 \in C([0,\infty) : \mathbb{R}^k)$ be such that $\omega_0(t) = z$ for some t. This implies that the open subset V of $C([0,\infty) : \mathbb{R}^k)$ with $\omega(0) = x$, and comprising all those ω such that $\omega(t) \in O$, has probability zero.]
2. Let the dimension $k = 1$, and $\mu(y) = 0$, $\sigma^2(y) = (1 + |y|^a)$, $y \in \mathbb{R}$. (i) Show that the diffusion is recurrent, whatever be $a \in \mathbb{R}$. (ii) Show that the diffusion is null recurrent if $a \leq 1$. (iii) Are there values of a for which the diffusion positive recurrent?
3. Let the dimension $k = 1$, and $\mu(y) = \theta \neq 0$, $\sigma^2(y) = (1 + |y|^a)$, $y \in \mathbb{R}$. (i) Show that the diffusion is transient for $a < 1$, and positive recurrent for $a > 1$. (iii) What happens if $a = 1$?
4. Let the dimension $k = 1$, and $\mu(y) = 0$, $\sigma^2(y) = e^{-|y|}$, $y \in \mathbb{R}$. Show that the diffusion is positive recurrent and compute its invariant distribution.
5. Classify the following diffusions as transient, null recurrent or positive recurrent.

   (i) $\mu(x) \equiv \mu \neq 0$, $\sigma^2(x) \equiv \sigma^2 > 0$.
   (ii) $\mu(x) \equiv 0$, $\sigma^2(x) \equiv \sigma^2 > 0$.
   (iii) $\mu(x) = \beta x$ $(\beta > 0)$, $\sigma^2(x) \equiv \sigma^2 > 0$.
   (iv) $\mu(x) = \beta x$ $(\beta < 0)$, $\sigma^2(x) \equiv \sigma^2 > 0$.

6. Consider a dimension $k \geq 2$. Let the drift vector $\mu(\cdot) = 0$, and the diffusion matrix given by $(1 + |y|^a)I_k$, $y \in \mathbb{R}^k$, for some constant a, I_k being the identity

matrix. (i) Show that, for all a, the diffusion is recurrent if $k = 2$ and transient if $k > 2$. (ii) Prove that if $a > 2$, the diffusion is positive recurrent for $k = 2$.

7. Let $k \geq 2$ be the dimension, and let $\mu(y) = \theta y$ for some $\theta > 0$, and let the diffusion matrix be $(1 + |y|^a)I_k$. (i) Check that for all $a < 2$, the diffusion is transient, whatever be the dimension $k \geq 2$. (ii) Show that for $a > 2$, the diffusion is recurrent for $k = 2$, and transient for $k > 2$. In this model the diffusion is transient for $k > 2$, whatever be a. (iii) Show that for $k = 2$, the diffusion is null recurrent if $a \geq 2$.
8. Next let $k \geq 1$, and consider the model in Exercise 7, but with $\theta < 0$. Show that the diffusion is positive recurrent for all values of a.
9. Assume that the coefficients of L are Hölder continuous and $((a_{ij}(x)))$ is positive definite for all x. Prove that $p(1; x, y)$ is bounded away from zero on every disc $\overline{B(x_0, \delta)}$. [*Hint:* Under the Dirichlet (or zero)-boundary condition, the fundamental solution $p_N(t; x, y)$ to the initial-boundary value problem for L on $\overline{B(0 : N)}$ is continuous in $t, x, y (t > 0)$, and for all $x, y \in B(0 : N)$. In particular, $\inf_{x \in \overline{B(x_0:\delta)}} p_N(1; x, y) = p_N(1; x_0, y)$, say, is positive if $\overline{B(x_0 : \delta)} \subset B(0 : N)$. Since, as $N \to \infty$, $p_N(1; x, y) \uparrow p(1; x, y)$, the transition probability of the diffusion at $t = 1$, it follows that $p(1; x, y) \geq p_N(1; x_0, y)$ for all $x \in \overline{B(x_0 : \delta)}$.]
10. Suppose $\{X(t) : t \geq 0\}$ is a positive recurrent diffusion satisfying the condition in Lemma 9(a). Prove that the distribution of the after-t process $X_t^+ = \{X(t + s) : s \geq 0\}$ under the distribution μ converges in variation distance to P_π as $t \to \infty$, where π is the invariant probability of the diffusion. [*Hint:* By Theorem 11.7, the distribution $\tilde{P}$, say, of $X_n^+ = \{X(n + m) : m = 0, 1, 2, \ldots\}$ converges in variation distance to $P_{\tilde{\pi}}$, the distribution of $\{X(n) : n = 0, 1, 2 \ldots\}$ under the invariant initial distribution π. In particular, $p(n; x, dy)$ converges in variation distance to π as $n \to \infty$. Use the Chapman-Kolmogorov equation $p(t; x, dy) = \int_{\mathbb{R}^k} p(n; x, dz)p(t - n; z, dy), t > n$, to argue that $p(t; x, dy)$ converges in variation distance to $\pi(dy)$ as $t \to \infty$. Finally, for an arbitrary Borel set $B \subset C([0, \infty) : \mathbb{R}^k)$, use $P_x(X_t^+ \in B) = \int_{\mathbb{R}^k} P_y(B)p(t; x, dy)$ to argue that $P_x(X_t^+ \in B) \to P_\pi(B)$, uniformly for all such Borel subsets B.]
11. Provide the details to verify each case of Example 1.
12. For the proof of part (a) of Theorem 11.7, prove that a sigma-finite invariant measure for the discrete skeleton $\{X(n) : n \geq 0\}$ is a sigma-finite invariant measure for the diffusion $\{X(t) : t \geq 0\}$. [*Hint:* Suppose π is a sigma-finite invariant measure of $\{X(n) : n \geq 0\}$. Choose any non-negative, bounded measurable function f on $\mathbb{R}^k$ such that $f = 0$ on $\mathbb{R}^k \backslash B$, where B is compact and $\pi(B) < \infty$. Consider $\{X(n) : n \geq 0\}$. One has $\int_{\mathbb{R}^k} \mathbb{E}_x(f(X(n))\pi(dx) = \int_{\mathbb{R}^k} f(x)\pi(dx), n = 0, 1, 2, \ldots$, by assumption. Also, by applying the above equality to $T_t f$, in place of f, one has $\int_{\mathbb{R}^k} T_{1+t} f d\pi = \int_{\mathbb{R}^k} T_t f d\pi$ for all $0 \leq t \leq 1$. Hence the blocks $\{\int_{\mathbb{R}^k} T_t f d\pi : n \leq t < n + 1\}$ are all identical. Thus, it follows that $\frac{1}{t} \int_{[0,t]} (\int_{\mathbb{R}^k} T_s f(x)\pi(dx))ds \to c(f) := \int_{[0,1]} (\int_{\mathbb{R}^k} T_s f(x)\pi(dx))ds$, as $t \to \infty$. Apply this to the function $T_t f$ to show that π is invariant under time shift of the diffusion.]

Chapter 12
Criteria for Explosion

The phenomena of explosion of a multidimensional nonsingular diffusion is considered in this chapter. In particular, conditions are provided that are sufficient for both explosion and nonexplosion. In one dimension, the conditions are also necessary.

Recall the simple Example 2 of Chapter 7 for a stark illustration of explosion of a diffusion having locally Lipschitzian coefficents. We derive in this chapter verifiable criteria for explosion (see Section 7.3). Throughout this chapter it is assumed, unless otherwise specified,[1]

$$\boldsymbol{\mu}(\cdot), \boldsymbol{\sigma}(\cdot) \text{satisfy Condition (A) of Chapter 11.} \tag{12.1}$$

First consider the case of one-dimensional diffusions. It is convenient to use Feller's scale and speed functions given by (11.22) and (11.23), respectively, given by

$$s(x) := \int_0^x e^{-I(0,y)} dy = \int_0^x \exp\{-\int_0^y \frac{2\mu(z)}{\sigma^2(z)}\, dz\} dy, \tag{12.2}$$

$$m(x) := \int_0^x \frac{2}{\sigma^2(y)} e^{I(0,y)} dy = \int_0^x \frac{2}{\sigma^2(y)} \exp\{\int_0^y \frac{2\mu(z)}{\sigma^2(z)} dz\} dy. \tag{12.3}$$

[1] The positivity condition is removed in the treatment of one-dimensional diffusions in Chapter 21.

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_12

Note that a diffusion $X^x, x \in \mathbb{R}$, having drift $\mu(\cdot)$ and diffusion coefficient $\sigma^2(\cdot)$ has the generator

$$L = \frac{1}{2}\sigma^2(x)\frac{d^2}{dx^2} + \mu(x)\frac{d}{dx} = \frac{d}{dm(x)}\,\frac{d}{ds(x)}\,. \tag{12.4}$$

By Proposition 8.7, for all $c < x < d$,

$$\begin{aligned} P\left(\tau_d^x < \tau_c^x\right) &= \frac{s(x) - s(c)}{s(d) - s(c)}\,, \quad \rho_{xd} = \frac{s(x) - s(-\infty)}{s(d) - s(-\infty)} \\ P(\tau_c^x < \tau_d^x) &= \frac{s(d) - s(x)}{s(d) - s(c)}\,, \quad \rho_{xc} = \frac{s(\infty) - s(x)}{s(\infty) - s(c)} \end{aligned} \tag{12.5}$$

with the convention $\infty/\infty = 1$. By Theorem 8.8, the diffusion is recurrent if and only if $s(-\infty) = -\infty$ and $s(\infty) = \infty$ (also see Exercise 2). Here ρ_{xy} is the probability that, starting at x, the diffusion ever reaches y. The following result shows that if the diffusion is nonrecurrent (i.e., if it is transient), then the diffusion $X^x(t)$ ($t \geq 0$) approaches an infinite limit $+\infty$ or $-\infty$ as t increases. In one dimension let us set $X^x(t) = \infty$ for $t \geq \zeta^x$ if $X^x(t) \to \infty$ as $t \uparrow \zeta^x < \infty$, where ζ^x is the *explosion time*. Similarly, set $X^x(t) = -\infty$ for $t \geq \zeta^x$ if $X^x(t) \to -\infty$ as $t \uparrow \zeta^x < \infty$.

Proposition 12.1

a. *If* $s(-\infty) = -\infty$ *and* $s(\infty) < \infty$, *then*

$$P(\inf_{t\geq 0} X^x(t) > -\infty \text{ and } \lim_{t\uparrow\zeta^x} X^x(t) = \infty) = 1. \tag{12.6}$$

b. *If* $s(-\infty) > -\infty$ *and* $s(\infty) = \infty$, *then*

$$P(\sup_{t\geq 0} X^x(t) < \infty \text{ and } \lim_{t\uparrow\zeta^x} X^x(t) = -\infty) = 1. \tag{12.7}$$

c. *If* $s(-\infty) > -\infty$ *and* $s(\infty) < \infty$, *then*

$$\begin{aligned} P(\inf_{t\geq 0} X^x(t) > -\infty \text{ and } \lim_{t\uparrow\zeta^x} X^x(t) = \infty) &= \frac{s(x) - s(-\infty)}{s(\infty) - s(-\infty)}\,, \\ P(\sup_{t\geq 0} X^x(t) < \infty \text{ and } \lim_{t\uparrow\zeta^x} X^x(t) = -\infty) &= \frac{s(\infty) - s(x)}{s(\infty) - s(-\infty)}\,. \end{aligned} \tag{12.8}$$

Proof (a) By (12.5), $\rho_{xd} = 1$ for all $x < d$, if $s(-\infty) = -\infty$. Letting $d \uparrow \infty$, one gets $P(\sup_{t\geq 0} X^x(t) = \infty) = 1$. If, also, $s(\infty) < \infty$, then by (12.5), $\rho_{xc} \downarrow 0$ as $c \downarrow -\infty (x > c)$. Hence $P(\inf_{t\geq 0} X^x(t) = -\infty) = 0$. Let $M > x$ be an integer. Then, letting N be an integer greater than M, and using the strong Markov property with τ_N^x,

$$
\begin{aligned}
&P(\lim_{t\uparrow\zeta^x} X^x(t) < M)\\
&\quad = P(\lim_{t\uparrow\zeta^x} X^N(t) < M)\\
&\quad = \rho_{NM} \equiv P(X^N(t) = M \text{ for some} t \geq 0) \to 0 \text{ as } N \uparrow \infty.
\end{aligned}
$$

Hence $P(\underline{\lim}_{t\uparrow\zeta^x} X^x(t) < M) = 0$ for every $M < \infty$, proving (12.6). The proof of (b) is entirely analogous.

(c) Let $x > c$. Then $P(X^x$ never reaches $c) \equiv 1-\rho_{xc} = 1-(s(\infty)-s(x))/(s(\infty)-s(c)) \uparrow (s(x)-s(-\infty))/(s(\infty)-s(-\infty))$(as $c \downarrow -\infty) = P(\inf_{t\geq 0} X^x(t) > -\infty)$. Similarly, $P(\sup_{t\geq 0} X^x(t) < \infty)$ is given by $(s(\infty) - s(x))/(s(\infty) - s(-\infty))$. To prove (12.8), it is then enough to show that $X^x(t)$ goes to either $+\infty$ or $-\infty$ as $t \uparrow \zeta_x$, outside a P-null set. For this, let M be a large integer such that $-M < x < M$. Let $N > M$. Then, by (12.5) applied to $c = -N$ and $d = N$, $\tau^x_{-N,N} := \tau^x_{-N} \wedge \tau^x_N < \infty$ a.s.. Using the strong Markov property with respect to $\tau^x_{-N,N} = \tau$, say one has

$$
\begin{aligned}
&P(|X^x(t)| > M \;\forall\; t > \tau^x_{-N,N}) =\\
&\quad = 1 - \mathbb{E}\{P(X^y \text{ reaches } [-M, M])_{y=X^x(\tau)}\}\\
&\quad = 1 - \mathbb{E}[\mathbf{1}_{[X^x(\tau)=-N]} P(X^{-N} \text{ reaches } -M) + \mathbf{1}_{[X^x(\tau)=N]} P(X^N \text{ reaches } M)]\\
&\quad = 1 - \{P(\tau^x_{-N} < \tau^x_N)\rho_{-N,-M} + P(\tau^x_N < \tau^x_{-N})\rho_{N,M}\} \longrightarrow 1
\end{aligned}
$$

since $\rho_{-N,-M} = (s(-N) - s(-\infty))/(s(-M) - s(-\infty))$ and $\rho_{N,M} = (s(\infty) - s(N))/(s(\infty) - s(M))$ both go to zero as $N \uparrow \infty$. Hence $P(\underline{\lim}_{t\uparrow\zeta^x} |X^x(t)| \geq M) = 1$. Now, let $M \uparrow \infty$ to get the desired result $P(\underline{\lim}_{t\uparrow\zeta^x} |X^x(t)| = \infty) = 1$, i.e., $P(\lim_{t\uparrow\zeta^x} |X^x(t)| = \infty) = 1$. ■

We are now ready to derive Feller's necessary and sufficient condition for explosion. To simplify notation, we will omit the superscript x in $\tau^x_z := \inf\{t \geq 0 : X^x(t) = z\}$ and write it as τ_z. Define

$$
\tau_{+\infty} := \lim_{n\uparrow\infty} \tau_n, \qquad \tau_{-\infty} := \lim_{n\uparrow\infty} \tau_{-n}, \tag{12.9}
$$

and note that by Proposition 12.1, if at least one of $s(-\infty)$, $s(\infty)$ is finite, then $\lim_{t\uparrow\zeta^x} X^x(t)$ exists and equals $+\infty$ or $-\infty$, a.s., and

$$
\begin{aligned}
\zeta^x &= \tau_{+\infty} \quad \text{on} \quad [\lim_{t\uparrow\zeta^x} X^x(t) = +\infty],\\
\zeta^x &= \tau_{-\infty} \quad \text{on} \quad [\lim_{t\uparrow\zeta^x} X^x(t) = -\infty].
\end{aligned}
\tag{12.10}
$$

Also, in the case $s(-\infty) = -\infty$ and $s(\infty) = \infty$, the diffusion is recurrent (Theorem 8.8) and, clearly, $\zeta^x = \infty$ a.s. for every x, as are $\tau_{+\infty}$ and $\tau_{-\infty}$. Hence,

explosion occurs (with positive probability) if and only if $P(\tau_{+\infty} < \infty) > 0$ or $P(\tau_{-\infty} < \infty) > 0$, for some x.

Theorem 12.2 (Feller's Test for Explosion) *Let $k = 1$, $I(z) := \int_0^z \frac{2\mu(v)}{\sigma^2(v)}\,dv$.*

a. *Suppose the integrals (i), (ii) below both diverge. Then $P(\zeta^x = \infty) = 1$ for all x.*

b. *If the integral (i) converges, then $P(\tau_{+\infty} < \infty) > 0$, and if the integral (ii) converges, then $P(\tau_{-\infty} < \infty) > 0$.*

$$\begin{aligned}
&\text{(i)} \int_0^\infty \exp\{-I(y)\}[\int_0^y (\exp\{I(v)\}/\sigma^2(v))dv)]dy, \\
&\text{(ii)} \int_{-\infty}^0 \exp\{-I(y)\}[\int_y^0 (\exp\{I(v)\}/\sigma^2(v))dv]dy.
\end{aligned} \tag{12.11}$$

Proof

(a) Suppose the integrals in (12.11) both diverge. We will construct a nonnegative function $\varphi(z)$ such that (1) $\varphi(z) \uparrow \infty$ as $0 \le z \uparrow \infty$ and $\varphi(z) \uparrow \infty$ as $z \downarrow -\infty$ $(z < 0)$, and (2) $A\varphi(z) \le \varphi(z)$ for all z. Given such a function, let $\psi(t, z) := e^{-t}\varphi(z)$. By Itô's lemma and Doob's optional stopping rule,

$$\begin{aligned}
&\mathbb{E}\psi(\tau_n \wedge \tau_{-n} \wedge t,\ X^x(\tau_n \wedge \tau_{-n} \wedge t)) - \varphi(x) \\
&= \mathbb{E}\int_0^{\tau_n\wedge\tau_{-n}\wedge t} e^{-s}[A\varphi(X^x(s)) - \varphi(X^x(s))]ds \le 0.
\end{aligned} \tag{12.12}$$

This means $\varphi(x) \ge \mathbb{E}\psi(\tau_n \wedge \tau_{-n} \wedge t, X^x(\tau_n \wedge \tau_{-n} \wedge t))$, so that

$$\varphi(x) \ge e^{-t}(\varphi(n) \wedge \varphi(-n))P(\tau_n \wedge \tau_{-n} \le t),$$
$$P(\tau_n \wedge \tau_{-n} \le t) \le e^t\varphi(x)\big/(\varphi(n) \wedge \varphi(-n)).$$

Letting $n \uparrow \infty$ in the last inequality, we get $P(\zeta^x \le t) = 0\ \forall t \ge 0$. This implies $P(\zeta^x < \infty) = 0$. To construct such a φ, write

$$\varphi_1(z) := \begin{cases} \int_0^z e^{-I(y)}m(y)dy = 2\int_0^z \exp\{-I(y)\}[\int_0^y \frac{\exp\{I(v)\}}{\sigma^2(v)}dv]dy & z \ge 0 \\ -\int_z^0 e^{-I(y)}m(y)dy = 2\int_z^0 \exp\{-I(y)\}[\int_y^0 \frac{\exp\{I(v)\}}{\sigma^2(v)}dv]dy, & z < 0, \end{cases} \tag{12.13}$$

and check that $L\varphi_1(z) = 1$ $(z \in \mathbb{R})$. Now define $\varphi(z) := 1 + \varphi_1(z)$, which satisfies all the requirements above.

(b) Assume first that the integral (i) in (12.11) converges. We will construct a function $\varphi(z)$ such that (1) $\varphi(z)$ increases to a finite positive limit as $z \uparrow \infty$,

$\varphi(0) = 0$, and (2) $L\varphi(z) \geq \varphi(z)\ \forall z > 0$. Given such a function, consider $\psi(t, y) := e^{-t}\varphi(z)$, and apply Itô's lemma to get, as in (12.12) but with $\tau = \tau_n \wedge \tau_0 \wedge t$ and with '≥ 0' in place of '≤ 0',

$$\begin{aligned}\varphi(x) &\leq \mathbb{E}\psi(\tau_n \wedge \tau_0 \wedge t, X^x(\tau_n \wedge \tau_0 \wedge t)) \\ &\leq \varphi(n)\left[P(\tau_n \leq \tau_0 \wedge t) + e^{-t}\right],\ x > 0,\end{aligned}$$

since $\varphi(0) = 0$. From this, one gets $P(\tau_n \leq \tau_0 \wedge t) + e^{-t} \geq \varphi(x)/\varphi(n)$, and letting $n \to \infty$, we have

$$e^{-t} + P(\tau_{+\infty} \leq \tau_0 \wedge t) \geq \frac{\varphi(x)}{\varphi(\infty)} > 0.$$

Now let $t \to \infty$ to derive the relation

$$P\ (\tau_{+\infty} < \infty) \geq \varphi(x)/\varphi(\infty) > 0. \tag{12.14}$$

It remains to construct φ. For this, let $\varphi_0(z) = 1$, and recursively define (note that $\varphi_1(z)$ is as given in (12.13) for $t \geq 0$)

$$\varphi_m(z) := 2\int_0^z \exp\{-I(y)\}[\int_0^y \frac{e^{I(v)}}{\sigma^2(v)}\varphi_{m-1}(v)dv]dy, \quad (z \in \mathbb{R}),\ m \geq 1. \tag{12.15}$$

Check that $L\varphi_m(z) = \varphi_{m-1}(z)$. Observe that $\varphi_m(z) \uparrow$ as $z \uparrow$ $(z \geq 0)$ and, as we will see by induction,

$$\varphi_m(z) \leq \frac{\varphi_1^m(z)}{m!} \qquad \forall\ m \geq 1,\ z \geq 0. \tag{12.16}$$

Assuming this for a given m, one has

$$\begin{aligned}\varphi_{m+1}(z) &\leq 2\int_0^z \exp\{-I(y)\}[\varphi_m(y)\int_0^y (\exp\{I(v)\}/\sigma^2(v))dv]dy \\ &\leq 2\int_0^z \exp\{-I(y)\}\frac{\varphi_1^m(y)}{m!}[\int_0^y (\exp\{I(v)\}/\sigma^2(v))dv]dy \\ &= \int_0^z \frac{d\varphi_1^{m+1}(y)}{(m+1)!} = \frac{\varphi_1^{(m+1)}(z)}{(m+1)!}\,,\end{aligned}$$

proving (12.16), since it clearly holds for $m = 1$. Now, by assumption $\varphi_1(z) < \varphi_1(\infty) < \infty$. Hence, there exists a positive integer r such that $\varphi_1^r(z)/r! < \frac{\varphi_1^r(\infty)}{r!} < 1\ \forall z \geq 0$. With this r define $\varphi(z) := \sum_{m=1}^r \varphi_m(z)$, so that $\varphi(z)$ increases to a finite limit as $z \uparrow \infty$, $\varphi(0) = 0$, and $L\varphi(z) = 1 + \sum_{m=1}^{r-1}\varphi_m(z) \geq \sum_{m=1}^r \varphi_m(z) \equiv \varphi(z) \geq 0$. This proves $P(\tau_{+\infty} < \infty) > 0$. The case, $P(\tau_{-\infty} <$

$\infty) > 0$ if the integral (ii) in (12.11) converges, is proved analogously (see Exercise 5).

■

Definition 12.1 In the case that $P(\zeta^x = \infty) = 1$ for all x, we say that the diffusion is *nonexplosive* or *conservative*.

Note that if the diffusion X on I is nonexplosive then $P_x(X(t) \in I) = 1$ for all $x \in I$.

To derive a criterion for explosion for a diffusion $\mathbf{X}^{\mathbf{x}}(t)$, $t \geq 0$, in dimension $k > 1$, the following notation will be used. Recall the notation of (12.1).

$$
\begin{aligned}
&\mathbf{D}(\mathbf{x}) = ((a_{ij}(\mathbf{x}))) := \boldsymbol{\sigma}(\mathbf{x})\boldsymbol{\sigma}'(\mathbf{x}), && A(\mathbf{x}) := \sum_{i,j} a_{ij}(\mathbf{x})x^{(i)}x^{(j)}/|\mathbf{x}|^2,\\
&C(\mathbf{x}) := \sum_i a_{ii}(\mathbf{x}), && B(\mathbf{x}) := 2\sum_i x^{(i)}\mu^{(i)}(\mathbf{x}),\\
&\overline{\beta}(r) := \max_{|\mathbf{x}|=r} \frac{B(\mathbf{x}) + C(\mathbf{x})}{A(\mathbf{x})} - 1, && \underline{\beta}(r) := \min_{|\mathbf{x}|=r} \frac{B(\mathbf{x}) + C(\mathbf{x})}{A(\mathbf{x})} - 1,\\
&\overline{\alpha}(r) := \max_{|\mathbf{x}|=r} A(\mathbf{x}), && \underline{\alpha}(r) := \min_{|\mathbf{x}|=r} A(\mathbf{x}),\\
&\overline{I}(r) := \int_{r_0}^{r} \frac{\overline{\beta}(u)}{u}\,du, && \underline{I}(r) := \int_{r_0}^{r} \frac{\underline{\beta}(u)}{u}\,du,
\end{aligned}
\tag{12.17}
$$

where $r_0 > 0$ is a given constant. It is simple to check that for every real valued twice continuously differentiable function F on $(0, \infty)$, for the function $\varphi(\mathbf{x}) := F(|\mathbf{x}|)$, one has (Exercise 4)

$$
2(A\varphi(\mathbf{x})) = A(\mathbf{x})F''(|\mathbf{x}|) + \frac{B(\mathbf{x}) + C(\mathbf{x}) - A(\mathbf{x})}{|\mathbf{x}|}F'(|\mathbf{x}|), \quad (|\mathbf{x}| > 0). \tag{12.18}
$$

Theorem 12.3 (Khaśminskii's Test for Explosion)

a. $P(\zeta < \infty) = 0$ *if* $\int_1^{\infty} e^{-\bar{I}(r)}\left(\int_1^r \frac{\exp\{\bar{I}(u)\}}{\bar{\alpha}(u)}\,du\right)dr = \infty,$

b. $P(\zeta < \infty) > 0$ *if* $\int_1^{\infty} e^{-\underline{I}(r)}\left(\int_1^r \frac{\exp\{\underline{I}(u)\}}{\underline{\alpha}(u)}\,du\right)dr < \infty.$

Proof

(a) Assume that the right side of (a) diverges. Fix $r_0 \in [0, |\mathbf{x}|)$. Define the following functions on (r_0, ∞),

$$\varphi_0(v) \equiv 1, \quad \varphi_n(v) = 2\int_{r_0}^{v} e^{-\bar{I}(r)}\Big(\int_{r_0}^{r} \frac{\exp\{\bar{I}(u)\}\varphi_{n-1}(u)}{\bar{\alpha}(u)}\,du\Big)dr, \quad n \geq 1. \tag{12.19}$$

Then the radial functions $\varphi_n(|\mathbf{y}|)$ on $(r_0 \leq |\mathbf{y}| < \infty)$ satisfy: $\mathbf{L}\varphi_n(|\mathbf{y}|) \leq \varphi_{n-1}(|\mathbf{y}|)$ (Exercise 4). Now, as in the proof of Theorem 12.2, apply Itô's lemma to the function

$$\varphi(t,\mathbf{y}) := \exp\{-t\}\varphi(|\mathbf{y}|), \qquad \text{where } \varphi(|\mathbf{y}|) = 1 + \varphi_1(|\mathbf{y}|),$$

extended to all of $\mathbb{R}^k$ as a twice continuously differentiable function. Then, writing $\tau_R := \inf\{t \geq 0 : |\mathbf{X}^{\mathbf{x}}(t)| = R\}$,

$$\mathbb{E}\varphi(\tau_{r_0} \wedge \tau_R \wedge t, \mathbf{X}^{\mathbf{x}}(\tau_{r_0} \wedge \tau_R \wedge t)) \leq \varphi(|\mathbf{x}|),$$

or

$$e^{-t}\varphi(R)P(\tau_R \leq \tau_{r_0} \wedge t) \leq \varphi(|\mathbf{x}|), \qquad P(\tau_R \leq \tau_{r_0} \wedge t) \leq e^t\,\frac{\varphi(|\mathbf{x}|)}{\varphi(R)}.$$

Letting $R \to \infty$, we get $P(\zeta \leq \tau_{r_0} \wedge t) = 0$ for all $t \geq 0$. Now let $r_0 \downarrow 0$ to get $P(\zeta < t) = 0\ \forall\, t$, since $\tau_{r_0} \to \infty$ a.s. as $r_0 \downarrow 0$. It follows that $P(\zeta < \infty) = 0$.

(b) This follows by replacing *upper bars* by *lower bars* in the definition (12.19) and noting that for the resulting functions, $\mathbf{L}\varphi_r(|\mathbf{y}|) \geq \varphi_{r-1}(|\mathbf{y}|)$. The rest of the proof follows the proof of the second half of part (a) of Theorem 12.2, with $\varphi(|y|) = \sum_{m=1}^{r}\varphi_m(|y|)$, where r is so chosen that $\varphi_1^r(\infty)/r! < 1$ (Exercise 4). ■

Exercises

1. Prove that no explosion occurs for a one-dimensional diffusion if $\mu(\cdot) \equiv 0$, and $\sigma^2(\cdot) > 0$ is locally Lipschitzian.
2. Check the following examples of drift and diffusion coefficients for recurrence, transience, explosion, or nonexplosion, where k is a nonnegative integer, β, σ are positive parameters.

   (a) $\mu(x) = -\beta x^{2k+1}, \sigma(x) = \sigma$.
   (b) $\mu(x) = -\beta x^{2k}, \sigma(x) = \sigma$.
   (c) $\mu(x) = -\beta x^{2k+1}, \sigma(x) = \sigma \cdot (1 + x^2)$.
   (d) $\mu(x) = -\beta x^{2k}, \sigma(x) = \sigma \cdot (1 + x^2)$.

3. Let $\sigma^2(\cdot)$ be bounded away from zero and infinity. Assume $\mu(x)$ is bounded on $x \leq 0$, and $\mu(x) \asymp x^c$ as $x \to \infty$ (i.e., there exist $0 < A_1 \leq A_2$ such that

$A_1 \le \mu(x)/x^c \le A_2$ for all sufficiently large x). Prove that explosion occurs if and only if $c > 1$.

4. (i) Derive (12.18).
   (ii) For the function φ_r in (12.19), prove that $L\varphi_r(|\mathbf{y}|) \le \varphi_{r-1}(|\mathbf{y}|)$.
   (iii) If φ_r are defined as in (12.19), but with upper bars replaced by lower bars, prove that $L\varphi_r(|\mathbf{y}|) \ge \varphi_{r-1}(|\mathbf{y}|)$.
   (iv) Write out the details of the proof of part (b) of Theorem 12.3.
   (v) Suppose that the functions $A(\mathbf{x})$ and $B(\mathbf{x}) + C(\mathbf{x})$ are radial, i.e., functions of $|\mathbf{x}|$. Prove that the criteria in Theorem 12.3 are then precise, i.e., they provide necessary and sufficient conditions for explosion.
5. Prove part (b) of Theorem 12.2 for the case when the second integral (ii) in (12.11) converges. [*Hint*: Relabel the Markov process by the transformation $x \to -x$, and then use the result for $\tau_{+\infty}$.]
6. Suppose $\sigma^2(z) > 0$ for all z, bounded away from zero and infinity for $z \le 0$, $\sigma(z) \asymp z^{1+\beta}$ and $\mu(z) \asymp z^{1+\alpha}$ as $z \to \infty$ ($\alpha > 0$, $\beta > 0$). Also assume $\mu(z)$ is bounded for $z \le 0$. Prove that explosion occurs if and only if $2\beta < \alpha$.
7. Let $\sigma^2(z) > 0$ for all z, and $\mu(z) < 0$ for $z > A$, $\mu(z) > 0$ for $z < -B$, where A and B are positive numbers. Show that the diffusion is nonexplosive.
8. (i) Let $k = 2$ and consider a diffusion $d\mathbf{X}(t) = \boldsymbol{\sigma}(\mathbf{X}(t))d\mathbf{B}(t)$, $\boldsymbol{\sigma}(\cdot)$ nonsingular. Show that there is no explosion if $D(\mathbf{x}) = \sigma(\mathbf{x})\sigma'(\mathbf{x})$ has equal eigenvalues for every $\mathbf{x}$. In particular, consider the case $\sigma(\mathbf{x}) = (1 + |\mathbf{x}|^a)I_2$ for some $a \in (-\infty, \infty)$. (ii) Let $k > 2$ and $\boldsymbol{\mu} = 0$, and suppose the eigenvalues of $D(\mathbf{x})$ are all equal to $\alpha(\mathbf{x})$, say, where $\alpha(\mathbf{x}) \ge d_1 + c_1|\mathbf{x}|^a$, for some constants $d_1 \ge 0$, $c_1 > 0$, $a > 2$. Show that the diffusion is explosive.
9. Prove that if $s(\infty) = \infty$, then $P(\tau_{+\infty} < \infty) = 0$. [*Hint*: Use Proposition 12.1 and Theorem 12.2.] Similarly, show that $P(\tau_{-\infty} < \infty) = 0$ if $s(-\infty) = -\infty$.

Chapter 13
Absorption, Reflection, and Other Transformations of Markov Processes

This chapter builds on the construction of unrestricted Markov processes on a measurable state space, by further constructing Markov processes by means of various transformations such as reflection, rotation, periodic embeddings, and so on. Much more generally, a special class of transformations of an unrestricted diffusion is shown to preserve the Markov property. These transformed processes are shown to be diffusions when the transformation is also sufficiently smooth.

We begin this chapter with the simple yet useful observation that a one-to-one measurable map of a state space $(S, \mathcal{S})$ of a Markov process $\{X_t : t \geq 0\}$ onto a state space $(S', \mathcal{S}')$ gives rise to a Markov process[1] $\{Y_t := \varphi(X_t), t \geq 0\}$ on $(S', \mathcal{S}')$. After all, this is merely a relabeling of the states. For example, if $\{X_t : t \geq 0\}$ is a diffusion on $\mathbb{R}$ with coefficients $\mu(\cdot)$ and $\sigma(\cdot)$, and φ is a twice continuously differentiable strictly increasing (or strictly decreasing) map on $\mathbb{R}$ onto (a, b) $(-\infty \leq a < b \leq \infty)$, then $\{Y_t : t \geq 0\}$ is said to be a *diffusion on* (a, b). By Itô's lemma,

$$\begin{aligned} dY_t = d\varphi(X_t) &= \{\varphi'(X_t)\mu(X_t) + \frac{1}{2}\sigma^2(X_t)\varphi''(X_t)\}dt + \varphi'(X_t)\sigma(X_t)dB_t \\ &= \widetilde{\mu}(Y_t)dt + \widetilde{\sigma}(Y_t)dB_t \qquad (t \geq 0), \qquad (13.1) \end{aligned}$$

[1] See Rogers and Pitman (1981) and references therein for some other approaches to conditions for the Markov property of functions of Markov processes.

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_13

where

$$\widetilde{\mu}(y) := (\varphi'\mu + \frac{1}{2}\sigma^2\varphi'')(\varphi^{-1}(y)), \quad \widetilde{\sigma}(y) := (\varphi'\sigma)(\varphi^{-1}(y)). \tag{13.2}$$

We will say $\{Y_t : t \geq 0\}$ is a *diffusion on* (a, b) *with coefficients* $\widetilde{\mu}(\cdot)$, $\widetilde{\sigma}(\cdot)$. In this manner, one can define diffusions on arbitrary nondegenerate open intervals (a, b). Conversely, given such a diffusion on an interval (a, b) (with one or both of a, b finite), one can construct an "equivalent" diffusion on $\mathbb{R}$ by a C^2-diffeomorphism φ on (a, b) onto $\mathbb{R}$. A point worth noting here is that a conservative diffusion on (a, b), with at least one of the end points a, b finite, cannot be defined for arbitrarily specified Lipschitzian coefficients. The coefficients $\widetilde{\mu}(\cdot)$ and $\widetilde{\sigma}(\cdot)$ of a conservative diffusion on (a, b) must be such that a C^2-diffeomorphism onto $\mathbb{R}$ will give rise to a conservative diffusion on $\mathbb{R}$. Following the line of proof of Feller's criteria for explosion (Theorem 12.2), one may show that a diffusion on (a, b), with coefficients $\mu(\cdot), \sigma(\cdot)$, is conservative if and only if, for some $x_0 \in (a, b)$ (Exercise 2),

$$\int_a^{x_0} m(x)ds(x) = -\infty, \qquad \int_{x_0}^b m(x)ds(x) = \infty, \tag{13.3}$$

where $s(x)$ and $m(x)$ are defined, for an arbitrarily fixed $x_0 \in (a, b)$, by

$$s(x) := \int_{x_0}^x \exp\{-I(y)\}\,dy, \qquad m(x) := \int_{x_0}^x \frac{2}{\sigma^2(y)} \exp\{I(y)\}dy,$$
$$I(x) := \int_{x_0}^x \frac{2\mu(y)}{\sigma^2(y)}\,dy. \tag{13.4}$$

Note that if the first integral in (13.3) is convergent, then $P(\tau_a < \infty) > 0$, where $\tau_a = \lim_{z\downarrow a} \tau_z$, $\tau_z := \inf\{t \geq 0 : Y_t = z\}$.

Although one can prove the assertion above as well as necessary and sufficient conditions for transience, null and positive recurrence on (a, b), by using the results in Chapter 8 and at the end of Chapter 11 on $S = (-\infty, \infty)$ by a twice continuously differentiable diffeomorphism $\varphi : (a, b) \to (-\infty, \infty)$, it is more useful to derive the criteria directly. The method is the same as in Chapters 8 and 11. In Exercise 2, one shows that the diffusion on (a, b) is recurrent if and only if

$$s(a) = -\infty, \quad s(b) = +\infty, \tag{13.5}$$

and it is positive recurrent if and only if, in addition to (13.5),

$$m(a) > -\infty, \quad m(b) < \infty. \tag{13.6}$$

Definition 13.1 (Feller's Boundary Classification) If the first integral in (13.3) is convergent, then 'a' is said to be *accessible*. Similarly, 'b' is *accessible* if the second integral in (13.3) converges. If a boundary is not accessible, then it is said to be *inaccessible*.

Example 1 (A Zero-Seeking Brownian Motion) Consider the process on $\mathbb{R}$ defined by

$$dZ(t) = -(1-\sigma^2)(1+Z^2(t))Z(t)dt + \sigma(1+Z^2(t))dB(t), \quad Z(0) = z.$$

Define $X(t) = \varphi(Z(t)), t \geq 0$, where $\varphi : \mathbb{R} \to (-\frac{\pi}{2}, \frac{\pi}{2})$ is given by $\varphi(z) = \tan^{-1}(z)$, $z \in (-\infty, \infty)$. Then X is constrained to the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$ and satisfies

$$dX(t) = -\tan(X(t))dt + \sigma dB(t).$$

States of $X(t)$ located away from the origin, symmetrically on either side, experience an increasingly large pull back toward the origin as a function of distance, becoming infinite at the boundaries. It is interesting to consider the longtime behavior of X in terms of the parameter $\sigma \geq 0$ (Exercise 3).

In a similar fashion as above, one can define diffusions on C^2-diffeomorphic images of $\mathbb{R}^k$ $(k > 1)$, such as the unit ball $B(\mathbf{0} : 1)$, or the positive orthant $(0, \infty)^k$, etc. Itô's lemma may be used to derive the coefficients of the diffusion on the image space in terms of those on $\mathbb{R}^k$, and vice versa.

13.1 Absorption

Let G be an open subset of $\mathbb{R}^k$ $(k \geq 1)$ on which are defined locally Lipschitzian coefficients $\boldsymbol{\mu}(\cdot)$, $\boldsymbol{\sigma}(\cdot)$ which may be extended to all of $\mathbb{R}^k$ as locally Lipschitzian so that the diffusion on $\mathbb{R}^k$ is nonexplosive. For a given initial state $\mathbf{x} \in G$, let $\mathbf{X}^{\mathbf{x}}$ be the diffusion on $\mathbb{R}^k$ with these extended coefficients. Then define the map

$$\overline{\mathbf{X}}_t^{\mathbf{x}} := \mathbf{X}^{\mathbf{x}}(t \wedge \tau^{\mathbf{x}}), \quad \tau(f) := \inf\{t \geq 0 : f(t) \in \partial G\}, \ f \in C([0, \infty) : \overline{G}),$$
$$\tau^{\mathbf{x}} := \tau(\mathbf{X}^{\mathbf{x}}), \qquad (\mathbf{x} \in \overline{G}). \tag{13.7}$$

The process $\overline{\mathbf{X}}^{\mathbf{x}}$ $(\mathbf{x} \in \overline{G})$ is said to be a *diffusion* on $\overline{G}$ having coefficients $\boldsymbol{\mu}(\cdot)$, $\boldsymbol{\sigma}(\cdot)$, *with absorption* on ∂G. Let $\{\mathcal{F}_t : t \geq 0\}$ denote the filtration with respect to which the Markov property of the unrestricted diffusion $\mathbf{X}^{\mathbf{x}}$ $(\mathbf{x} \in \mathbb{R}^k)$ is defined.

Proposition 13.1 $\overline{\mathbf{X}}^{\mathbf{x}}$ $(\mathbf{x} \in \overline{G})$ *is a Markov process on the state space* $\overline{G}$, *with the transition probability* $\overline{p}(t; \mathbf{x}, d\mathbf{y})$ *defined by*

$$\overline{p}(t; \mathbf{x}, d\mathbf{y}) := \delta_{\mathbf{x}}(d\mathbf{y}) \quad \forall\, \mathbf{x} \in \partial G, \ t \geq 0,$$

$$\overline{p}(t; \mathbf{x}, B) := \begin{cases} P(\tau^{\mathbf{x}} > t, \mathbf{X}_t^{\mathbf{x}} \in B) \ \mathbf{x} \in G, B \in \mathcal{B}(G), \\ P(\tau^{\mathbf{x}} \leq t, \mathbf{X}_{\tau^{\mathbf{x}}}^{\mathbf{x}} \in B) \ \mathbf{x} \in G, B \in \mathcal{B}(\partial G). \end{cases} \tag{13.8}$$

Proof If $\mathbf{x} \in \partial G$, then $\overline{\mathbf{X}}_t^{\mathbf{x}} = \mathbf{x} \forall\ t \geq 0$ and it is clearly (trivially) Markov with the transition probability $\delta_{\mathbf{x}}(d\mathbf{y})$. Suppose $\mathbf{x} \in G$, B Borel subset of G. Then, writing $(\mathbf{X}_s^{\mathbf{x}})^+$ for the after-s process $\{\mathbf{X}_{s+t'}^{\mathbf{x}} : t' \geq 0\}$, one has

$$\begin{aligned} P(\overline{\mathbf{X}}_{s+t}^{\mathbf{x}} \in B \mid \mathcal{F}_s) &= P(\mathbf{X}_{s+t}^{\mathbf{x}} \in B,\ \tau^{\mathbf{x}} > s+t \mid \mathcal{F}_s) \\ &= P(\mathbf{X}_{s+t}^{\mathbf{x}} \in B, \tau^{\mathbf{x}} > s, \tau((\mathbf{X}_s^{\mathbf{x}})^+) > t \mid \mathcal{F}_s) \\ &= \mathbf{1}_{\{\tau^{\mathbf{x}} > s\}}[P(\mathbf{X}_t^{\mathbf{y}} \in B,\ \tau^{\mathbf{y}} > t)]_{\mathbf{y}=\mathbf{X}_s^{\mathbf{x}}} \quad \text{(strong Markov property)} \\ &= [P(\mathbf{X}_t^{\mathbf{y}} \in B,\ \tau^{\mathbf{y}} > t)]_{\mathbf{y}=\overline{\mathbf{X}}_s^{\mathbf{x}}} \equiv \overline{p}(t; \overline{\mathbf{X}}_s^{\mathbf{x}}, B), \end{aligned}$$

since $\overline{\mathbf{X}}_s^{\mathbf{x}} = \mathbf{X}_s^{\mathbf{x}}$ if $\tau^x > s$, and $\overline{p}(t; \overline{\mathbf{X}}_s^{\mathbf{x}}, B)$ is zero on $[\tau^x \leq s](B \subset G)$.

Now, let $B \subseteq \partial G$. Then

$$\begin{aligned} P(\overline{\mathbf{X}}_{s+t}^{\mathbf{x}} \in B \mid \mathcal{F}_s) &= P(\mathbf{X}_{\tau^{\mathbf{x}}}^{\mathbf{x}} \in B,\ \tau^{\mathbf{x}} \leq s+t \mid \mathcal{F}_s) \\ &= P(\mathbf{X}_{\tau^{\mathbf{x}}}^{\mathbf{x}} \in B,\ \tau^{\mathbf{x}} \leq s \mid \mathcal{F}_s) + P(\mathbf{X}_{\tau^{\mathbf{x}}}^{\mathbf{x}} \in B,\ s < \tau^{\mathbf{x}} \leq s+t \mid \mathcal{F}_s) \\ &= \mathbf{1}_{[\tau^{\mathbf{x}} \leq s] \cap [\mathbf{X}_{\tau^{\mathbf{x}}}^{\mathbf{x}} \in B]} + \mathbf{1}_{[\tau^{\mathbf{x}} > s]} P(\tau((\mathbf{X}_s^{\mathbf{x}})^+) \leq t,\ (\mathbf{X}_s^{\mathbf{x}})^+_{\tau((\mathbf{X}_s^{\mathbf{x}})^+)} \in B \mid \mathcal{F}_s) \\ &= \mathbf{1}_{[\tau^{\mathbf{x}} \leq s]} \overline{p}(t; \overline{\mathbf{X}}_s^{\mathbf{x}}, B) + \mathbf{1}_{[\tau^{\mathbf{x}} > s]}[P(\tau^{\mathbf{y}} \leq t, \mathbf{X}_{\tau^{\mathbf{y}}}^{\mathbf{y}} \in B)]_{\mathbf{y}=\mathbf{X}_s^{\mathbf{x}}} \\ &= \mathbf{1}_{[\tau^{\mathbf{x}} \leq s]} \overline{p}(t; \overline{\mathbf{X}}_s^{\mathbf{x}}, B) + \mathbf{1}_{[\tau^{\mathbf{x}} > s]} \overline{p}(t; \overline{\mathbf{X}}_s^{\mathbf{x}}, B) \\ &= \overline{p}(t; \overline{\mathbf{X}}_s^{\mathbf{x}}, B). \end{aligned}$$

■

Example 2 (Brownian Motion on a Half-Line with Absorption at Zero) Let $\{B_t^x : t \geq 0\}$ be a Brownian motion on $\mathbb{R} = (-\infty, \infty)$ with $B_0^x = x$, where $x < 0$. Let τ be as above, with $G = (-\infty, 0)$, $\partial G = \{0\}$. Consider Brownian motion W, reflected at $0 : W_t = B_t^x$ for $t \leq \tau^x$ and $W_t = -B_t^x$ for $t > \tau^x$. By the strong Markov property, W has the same distribution as B^x. Also $\tau^x \equiv \tau(B^x) = \tau(W)$. Therefore, for all $c < 0$,

$$\begin{aligned} \overline{p}(t; x, (-\infty, c]) &= P(B_t^x \leq c, \tau^x > t) \\ &= P(B_t^x \leq c) - P(B_t^x \leq c, \tau^x \leq t) \end{aligned} \tag{13.9}$$

Now,

$$
\begin{aligned}
P(B_t^x \le c, \tau(B^x) \le t) &= P(W_t \le c, \tau(W) \le t) \\
&= P(B_t^x \ge -c, \tau(W) \le t) = P(B_t^x \ge -c, \tau(B^x) \le t) \\
&= P(B_t^x \ge -c) \qquad (\text{Since`} B_t^x \ge -c' \implies \text{`}\tau(B^x) \le t'). \quad (13.10)
\end{aligned}
$$

Using (13.10) in (13.9), one obtains

$$\overline{p}(t; x, (-\infty, c]) = P(B_t^x \le c) - P(B_t^x \ge -c). \tag{13.11}$$

Also,

$$\overline{p}(t; x, \{0\}) = P(\tau^x \le t). \tag{13.12}$$

Thus, on $(-\infty, 0)$, $\overline{p}(t; x, dy)$ has the density (obtained by differentiating (13.11) with respect to c, and then setting $c = y$)

$$q_1(t; x, y) := p(t; x, y) - p(t; x, -y) = p(t; x, y) - p(t; -x, y) \quad (x < 0, y < 0), \tag{13.13}$$

where $p(t; x, y)$ is the transition probability density of a standard Brownian motion

$$p(t; x, y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}}.$$

Hence,

$$q_1(t; x, y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}} - \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y+x)^2}{2t}} \quad (x < 0, y < 0). \tag{13.14}$$

Note that the discrete component of $\overline{p}(t; x, dy)$ puts mass $\overline{p}(t; x, \{0\}) = P(\tau^x \le t)$ at $\{0\}$. One may write this probability explicitly as

$$\overline{p}(t; x, \{0\}) = 1 - \int_{(-\infty,0)} q_1(t; x, y) dy. \tag{13.15}$$

The transition probability of Brownian motion on $[0, \infty)$ with absorption at '0' may be obtained from that on $(-\infty, 0]$ by making the transformation $x \to -x$ on the state space $(-\infty, 0]$ to $[0, \infty)$. In this case, the density component of the transition probability $p(t; x, dy)$ ($x \in (0, \infty)$) is obtained from (13.14) as

$$
\begin{aligned}
q_2(t; x, y) = q_1(t; -x, -y) &= p(t; -x, -y) - p(t; -x, y) \\
&= p(t; x, y) - p(t; x, -y) \\
&= p(t; x, y) - p(t; -x, y), \qquad (x > 0, y > 0) \quad (13.16)
\end{aligned}
$$

13.2 General One-Dimensional Diffusions on Half-Line with Absorption at Zero

We first consider locally Lipschitz coefficients $\mu(\cdot), \sigma(\cdot)$ on $[0, \infty)$ with $\mu(0) = 0$ (and $\sigma(y) \neq 0$, $y \in [0, \infty)$). Extend these coefficients to all of $\mathbb{R}$ as follows:

$$\mu(-x) = -\mu(x), \quad \sigma(-x) = \sigma(x) \qquad (x > 0). \tag{13.17}$$

Then the unrestricted diffusion $\{X_t : t \geq 0\}$ on $(-\infty, \infty)$ satisfies the Itô equation with the same coefficients, namely, $\mu(\cdot), \sigma(\cdot)$, as does $\{Y_t := -X_t, t \geq 0\}$. In other words, they have the same law, starting from the same initial point. In particular, they have the same transition probability. One may now repeat the arguments in Example 2 verbatim, with Brownian motion $\{B_t : t \geq 0\}$ replaced by $\{X_t : t \geq 0\}$. In particular, assuming that X_t has a transition probability density $p(t; x, y)$ that satisfies the backward equations, the transition probability $\overline{p}(t; x, dy)$ of the process absorbed at 0, $x \in (-\infty, 0)$, has the density component

$$q(t; x, y) = p(t; x, y) - p(t; x, -y) = p(t; x, y) - p(t; -x, y), \quad (x < 0, y < 0). \tag{13.18}$$

Similarly (or using relabeling $x \to -x$), the density component of the transition probability $\overline{p}$ for the diffusion on $[0, \infty)$, with absorption at 0, satisfies the same equation as (13.18), but for $x > 0$, $y > 0$.

Remark 13.1 In the case $\mu(x) \neq 0$ in the specification of the drift on $(0, \infty)$, one may still extend $\mu(\cdot)$ and $\sigma(\cdot)$ to $(-\infty, 0)$ by (13.17), and one may take $\mu(0)$ to be arbitrary, say $\mu(0)$. As will be seen in Chapter 21, the construction of Feller's one-dimensional diffusion[2] is determined by the scale function $s(\cdot)$ and speed function $m(\cdot)$, which are not affected by changing the values of $\mu(\cdot)$ at one point. Thus, the formula (13.18) still holds even if $\mu(0-) \neq 0$.

Remark 13.2 By (13.18), one has

$$\lim_{x \uparrow 0} q(t; x, y) = 0 \qquad (t > 0, y < 0), \tag{13.19}$$

and

$$\lim_{y \uparrow 0} q(t; x, y) = 0 \qquad (t > 0, x < 0). \tag{13.20}$$

Also, $q(t; x, y)$ satisfies

$$\frac{\partial q(t; x, y)}{\partial t} = (Aq)(t; x, y) \qquad (t > 0,\ x < 0,\ y < 0), \tag{13.21}$$

[2] See Chapter 21, or Mandl (1968).

where $A = \mu(x)\, d/dx + \frac{1}{2}\sigma^2(x) d^2/dx^2$. It follows that $q(t; x, y)$ is the fundamental solution to the *initial-value* (or, Cauchy) problem:

$$\begin{aligned}
\frac{\partial u(t,x)}{\partial t} &= Au(t,x) \qquad (t > 0,\ x \in (-\infty, 0)),\\
\lim_{x \uparrow 0} u(t,x) &= 0 \qquad (t > 0),\\
\lim_{t \downarrow 0} u(t,x) &= f(x),
\end{aligned} \tag{13.22}$$

where f is a bounded uniformly continuous function on $(-\infty, 0)$ satisfying the boundary condition $f(0+) = 0$, and in this case, the solution to (13.22) is given by

$$u(t,x) = \int_{(-\infty,0)} f(y) q(t; x, y) dy, \qquad (t > 0, x < 0). \tag{13.23}$$

More generally, it may be shown that if G is an open set with a smooth boundary (e.g., $G = \{\mathbf{x} \in \mathbb{R}^k : |\mathbf{x}| < 1\}$) and

$$A = \sum_i \mu^{(i)}(\mathbf{x})\, \partial/\partial x^{(i)} + \frac{1}{2} \sum_{i,j} a_{ij}(\mathbf{x}) \partial^2/\partial x^{(i)} \partial x^{(j)},$$

with $\boldsymbol{\mu}(\cdot)$, $\mathbf{a}(\cdot) = \boldsymbol{\sigma}(\cdot)\boldsymbol{\sigma}'(\cdot)$ bounded and Lipschitzian and the eigenvalues of $\mathbf{a}(\cdot)$ being bounded away from zero and infinity, then the density component $q(t; \mathbf{x}, \mathbf{y})$ of the distribution $\overline{p}(t; \mathbf{x}, dy)$ $(\mathbf{x} \in \overline{G})$ is the fundamental solution to the initial boundary value problem[3]

$$\begin{aligned}
\frac{\partial u(t,\mathbf{x})}{\partial t} &= Au(t,\mathbf{x}) \qquad (t > 0, \mathbf{x} \in G),\\
\lim_{\substack{\mathbf{x} \to \partial G \\ (\mathbf{x} \in G)}} u(t,\mathbf{x}) &= 0 \qquad (t > 0),\\
\lim_{t \downarrow 0} u(t,\mathbf{x}) &= f(\mathbf{x}) \qquad (\mathbf{x} \in G),
\end{aligned} \tag{13.24}$$

for bounded uniformly continuous f on G satisfying the boundary condition

$$\lim_{\substack{\mathbf{x} \to \mathbf{b} \\ (\mathbf{x} \in G)}} f(\mathbf{x}) = 0 \qquad \forall\ b \in \partial G. \tag{13.25}$$

Let the diffusion on $\overline{G}$, with absorption on ∂G, be denoted by $\overline{X}^x$, as in Proposition 13.1. Note that if f can be extended to a twice continuously differentiable function as a bounded function with continuous and bounded derivatives, and if

[3] See Friedman (1964).

f satisfies the boundary condition (13.25), then by Itô's lemma and the optional stopping rule (noting that $\overline{\mathbf{X}}_t^x = \mathbf{X}^x_{\tau^x \wedge t}$), the function $u(t,x) := \mathbb{E} f(\overline{\mathbf{X}}_t^{\mathbf{x}})$ satisfies

$$\begin{aligned} u(t,\mathbf{x}) - f(\mathbf{x}) &\equiv \mathbb{E} f(\overline{\mathbf{X}}_t^{\mathbf{x}}) - f(\mathbf{x}) \\ &= \mathbb{E} \int_0^{\tau \wedge t} (Af)(\mathbf{X}_s^{\mathbf{x}}) ds = \mathbb{E} \int_0^t (Af)(\mathbf{X}_{s\wedge\tau}^{\mathbf{x}}) ds \\ &= \mathbb{E} \int_0^t (Af)(\overline{\mathbf{X}}_s^{\mathbf{x}}) ds = \int_0^t (\overline{T}_s Af)(x) ds, \end{aligned} \tag{13.26}$$

where $\overline{T}_s g(\mathbf{x}) := \int_{\overline{G}} g(\mathbf{y}) \overline{p}(s; \mathbf{x}, d\mathbf{y}))$, so that, using the fact that $\overline{T}_t$ and its generator A (restricted to functions f that vanish on ∂G) commute,

$$\frac{\partial u(t,\mathbf{x})}{\partial t} = \overline{T}_t Af(\mathbf{x}) = A\overline{T}_t f(\mathbf{x}) \equiv Au(t,\mathbf{x}) \qquad (t > 0, \mathbf{x} \in G),$$

yielding the first equation in (13.24). The third equation is clear from (13.26). For the middle equation, use the fact that if $\mathbf{x}$ $(\in G) \to \mathbf{b} \in \partial G$, then, for any given $t > 0$, $\tau^{\mathbf{x}} \to 0$, a.s., $\tau^{\mathbf{x}} \wedge t \to 0$ a.s., $\mathbf{X}^{\mathbf{x}}_{\tau^{\mathbf{x}} \wedge t} \to \mathbf{b}$ a.s., and, therefore, $u(t,\mathbf{x}) = \mathbb{E} f(\mathbf{X}^{\mathbf{x}}_{\tau^{\mathbf{x}} \wedge t}) \to f(\mathbf{b}) = 0$.

Example 3 (k-Dimensional Brownian Motion on a Half-Space with Absorption at the Boundary) Let $\mathbf{B}^{\mathbf{x}}$ be a Brownian motion on $\mathbb{R}^k$ $(k > 1)$, $\mathbf{x} \in [0,\infty) \times \mathbb{R}^{k-1} = \{\mathbf{y} = (y^{(1)}, \ldots, y^{(k)}) : y^{(1)} \geq 0\}$. Define $\overline{\mathbf{B}}_t^{\mathbf{x}} := B^{\mathbf{x}}_{t \wedge \tau^{\mathbf{x}}}$, where $\tau^{\mathbf{x}} := \inf\{t \geq 0 : (B_t^{\mathbf{x}})^{(1)} = 0\} = \tau^{x^{(1)}}$, say. The transition probability of $\overline{\mathbf{B}}_t^{\mathbf{x}}$ has the *density component*

$$\overline{p}(t; \mathbf{x}, \mathbf{y}) = q_2(t; x^{(1)}, y^{(1)}) \prod_{j=2}^{k} p(t; x^{(j)}, y^{(j)}) \quad \mathbf{x}, \mathbf{y} \in (0,\infty) \times \mathbb{R}^{k-1},$$

where q_2 is as in (13.16) and $p(t; x, y)$ is the transition probability density of a one-dimensional standard Brownian motion. The *singular part* of $\overline{p}(t; \mathbf{x}, d\mathbf{y})$, for $\mathbf{x} \in (0,\infty) \times \mathbb{R}^{k-1}$ is given by the following, where we write $\mathbf{x}_1 = (x^{(2)}, \ldots, x^{(k)})$ and $\tau^{x^{(1)}}$ as the first passage time to the boundary, starting at $(x^{(1)}, , \mathbf{x}_1)$,

$$\begin{aligned} &\overline{p}(t; \mathbf{x}, \{0\} \times C) \\ &\quad = P(\tau^{x^{(1)}} \leq t,\ \mathbf{B}^{\mathbf{x}}_{\tau^{x^{(1)}}} \in C) \\ &\quad = \int_0^t g_{x^{(1)}}(s) \Phi_{k-1}(s^{-\frac{1}{2}}(C - \mathbf{x}_1)) ds, \qquad C \in \mathcal{B}(\mathbb{R}^{k-1}), \end{aligned} \tag{13.27}$$

where $g_x(\cdot)$ is the first passage time density (to 0) of a one-dimensional standard Brownian motion starting at $x^{(1)} > 0$. This can be obtained from (13.15) on differentiation with respect to t but is given more explicitly in (16.6) of Bhattacharya

and Waymire (2021). Also see Example 1, Chapter 5. Here, Φ_{k-1} is the standard $(k-1)$-dimensional normal distribution (with mean vector zero and identity dispersion matrix). In the case $k = 2$, this distribution can be expressed in closed form (Exercise 5).

Example 4 (Markov Property for Functions of a Markov Process) We next consider the problem of constructing Markov processes Y_t, $t \geq 0$, which are functions of a Markov process X_t, $t \geq 0$, say, $Y_t = \varphi(X_t)$, where φ is not one-to-one, and thus Y_t may be viewed as a *reduction* of X_t. Here is a preliminary result.

Proposition 13.2 *Let $\{X_t^x, t \geq 0\}$, $x \in S$, be a time-homogeneous Markov process on a measurable state space $(S, \mathcal{S})$, having transition probability $p(t; x, dy)$ and semigroup of transition operators T_t, $t \geq 0$. Suppose φ is a measurable map on $(S, \mathcal{S})$ onto $(S', \mathcal{S}')$. If $T_t(f \circ \varphi)$ is a (measurable) function of φ for every bounded measurable f on S', then $Y_t^y := \varphi(X_t^x)$, $t \geq 0$, is a time-homogeneous Markov process on $(S', \mathcal{S}')$, with $y = \varphi(x)$ $(x \in S)$, having the transition probability $q(t; y, C) := p(t; x, \varphi^{-1}(C))$ $(x \in \varphi^{-1}(\{y\}))$.*

Proof Suppose X_t^x, $t \geq 0$, has the Markov property with respect to the filtration $\{\mathcal{F}_t, t \geq 0\}$. Then, using this Markov property, for a given $y \in S'$, and $x \in \varphi^{-1}(\{y\})$, one has for $C \in \mathcal{S}'$,

$$\begin{aligned} P(Y_{s+t}^y \in C \mid \mathcal{F}_s) &= P(\varphi(X_{s+t}^x) \in C \mid \mathcal{F}_s) \\ &= P(X_{s+t}^x \in \varphi^{-1}(C) \mid \mathcal{F}_s) = p(t; X_s^x, \varphi^{-1}(C)) = T_t(f \circ \varphi)(X_s^x), \end{aligned}$$

where $f = \mathbf{1}_C$. By assumption, $T_t(f \circ \varphi)(z) = v(t, \varphi(z))$ for a measurable function $y \to v(t, y)$. Hence,

$$P(Y_{s+t}^y \in C \mid \mathcal{F}_s) = v(t, \varphi(X_s^x)) = v(t, Y_s^y),$$

proving the Markov property of $\{Y_t^y, t \geq 0\}$ with transition probability $q(t; y, C) = p(t; x, \varphi^{-1}(C))$. ■

Since it is generally not possible to calculate $p(t; x, dy)$ explicitly (in order to check the hypothesis of Proposition 13.2), the next result can often be used to provide important examples.

Consider a *group $\mathcal{G}$ of transformations g* on S such that

$$p(t; g^{-1}x, g^{-1}B) = p(t; x, B) \qquad \forall\, t > 0, x \in S, B \in \mathcal{S}. \tag{13.28}$$

Here, each g is a bimeasurable map on S onto S. We say a measurable function φ on $(S, \mathcal{S})$ onto $(S', \mathcal{S}')$ is a *maximal invariant* if (1) $\varphi \circ g = \varphi \forall\, g \in \mathcal{G}$ (*invariance*) and (2) if ψ is an invariant function: $(S, \mathcal{S}) \to (S'', \mathcal{S}'')$, i.e., $\psi \circ g = \psi\ \forall\, g \in \mathcal{G}$, then there exists a measurable function $\gamma : S' \to S''$ such that $\psi = \gamma \circ \varphi$ (*maximality*). One may, for most purposes, think of S' as the *space of orbits* $\{g_x : g \in \mathcal{G}\}$, $x \in S$.

Theorem 13.3 *Under the hypothesis (13.28), $Y_t := \varphi(X_t)$, $t \geq 0$, is a Markov process, where φ is a maximal invariant under $\mathcal{G}$.*

Proof For all $C \in \mathcal{S}'$, $y \in S'$, $x \in \varphi^{-1}(\{y\})$, so that $Y_t^y = \varphi(X_t^x)$, one has

$$P(Y_{s+t}^y \in C \mid \mathcal{F}_s) = P(X_{s+t}^x \in \varphi^{-1}(C) \mid \mathcal{F}_s) = p(t; X_s^x, \varphi^{-1}(C)). \tag{13.29}$$

Now by assumption (13.28) and invariance of φ one has $\forall\, g \in \mathcal{G}$,

$$\begin{aligned} p(t; x, \varphi^{-1}(C)) &= p(t; g^{-1}x, g^{-1}\varphi^{-1}(C)) = p(t; g^{-1}x, (\varphi \circ g)^{-1}(C)) \\ &= p(t; g^{-1}x, \varphi^{-1}(C)), \end{aligned}$$

which shows that $x \to p(t; x, \varphi^{-1}(C))$ is invariant under $\mathcal{G}$. Hence, because of the maximality of φ, $p(t; x, \varphi^{-1}(C)) = q(t; \varphi(x), C)$, say, for some function q such that $y' \to q(t; y', C)$ is measurable on $(S', \mathcal{S}')$. Hence, by (13.29),

$$P\left(Y_{s+t}^y \in C \mid \mathcal{F}_s\right) = q(t; \varphi(X_s^x), C) = q(t; Y_s^y, C).$$

■

We will consider several applications of Proposition 13.2 and Theorem 13.3. The following immediate corollary allows one to verify the hypothesis of Theorem 13.3 in the case of diffusions.

Corollary 13.4 *If the generator A of a diffusion $\{X_t : t \geq 0\}$ is invariant under a group of transformations, and φ is the maximal invariant, then $\{Y_t = \varphi(X_t) : t \geq 0\}$ is a Markov process.*

Example 5 (General Diffusions with a Markovian Radial Component) Consider a diffusion on $\mathbb{R}^k$ $(k > 1)$ governed by the Itô equation

$$d\mathbf{X}_t = \boldsymbol{\mu}(\mathbf{X}_t)dt + \boldsymbol{\sigma}(\mathbf{X}_t)d\mathbf{B}_t, \quad t > 0. \tag{13.30}$$

To determine conditions under which the norm $|\mathbf{X}_t|$, $t \geq 0$, is a Markov process, one may look for conditions under which the generator A maps (smooth) radial functions to radial functions, i.e., smooth functions of the form, $h(\mathbf{x}) = F(|\mathbf{x}|)$. In anticipation of such conditions, let $d(\mathbf{x}) := \Sigma a_{ij}(\mathbf{x})x^{(i)}x^{(j)}/|\mathbf{x}|^2$ $(\mathbf{x} \neq \mathbf{0})$, $C(\mathbf{x}) := \Sigma a_{ii}(\mathbf{x})$, $B(\mathbf{x}) := 2\mathbf{x}\cdot\boldsymbol{\mu}(\mathbf{x}) \equiv 2\Sigma x^{(i)}\mu^{(i)}(\mathbf{x})$. Assume that the functions $d(\mathbf{x})$ and $B(\mathbf{x})+C(\mathbf{x})$ are radial. Assume also that $\boldsymbol{\sigma}(\cdot)$ is nonsingular, and write

$$d(\mathbf{x}) = \alpha(|\mathbf{x}|), \quad \beta(|\mathbf{x}|) = \frac{B(\mathbf{x}) + C(\mathbf{x}) - d(\mathbf{x})}{d(\mathbf{x})}, \quad (\mathbf{x} \neq 0). \tag{13.31}$$

If F is a twice continuously differentiable bounded function on $(0, \infty)$ with bounded derivatives, then for the radial function $h(\mathbf{x}) := F(|\mathbf{x}|)$ on $\mathbb{R}^k\backslash\{0\}$, one has (see Lemma 8, Chapter 11),

$$Ah(\mathbf{x}) = \frac{1}{2}d(\mathbf{x})F''(|\mathbf{x}|) + \frac{B(\mathbf{x}) + C(\mathbf{x}) - d(\mathbf{x})}{2|\mathbf{x}|}F'(|\mathbf{x}|) \quad (\mathbf{x} \neq 0), \tag{13.32}$$

so that A transforms radial functions to radial functions. It follows from Proposition 13.2 that $\mathbb{R}_t := |\mathbf{X}_t|$, $t \geq 0$, is a diffusion on $(0, \infty)$ with generator

$$A^R := \frac{1}{2}\alpha(r)[\frac{d^2}{dr^2} + \frac{\beta(r)}{r}\frac{d}{dr}]. \tag{13.33}$$

(see Exercise 6), with inaccessible boundaries 0 and ∞, or, equivalently, $\log R_t$, $t \geq 0$, is a conservative diffusion on $(-\infty, \infty)$. In this example, the group G in Corollary 13.4 is the group of rotations, i.e., the orthogonal group.

Example 6 (Bessel Processes) Let $k \geq 2$, $\{\mathbf{B}_t, t \geq 0\}$ be a standard Brownian motion on $\mathbb{R}^k$. As a special case of the previous example, $R_t := |\mathbf{B}_t|$, $t \geq 0$, is a diffusion on $(0, \infty)$ with generator

$$A^R := \frac{1}{2}\frac{d^2}{dr^2} + \frac{k-1}{2r}\frac{d}{dr}, \tag{13.34}$$

with inaccessible boundaries 0 and ∞.

Example 7 (Squared Bessel Process) Let $k \geq 2$ and $R_t^2 = |\mathbf{B}_t|^2$, $t \geq 0$ where $\mathbf{B}$ is standard Brownian motion started at $\mathbf{0}$. From the preceding example, it follows that $R_t^2 = |B_t|^2$, $t \geq 0$, is a diffusion on $(0, \infty)$, since $x \to x^2$ is a diffeomorphism. Here is an alternative description of the process. By Itô's lemma, one has

$$dR_t^2 = kdt + 2\mathbf{B}_t \cdot d\mathbf{B}_t.$$

Now observe, noting also that 0 is an inaccessible boundary for R, the process $\tilde{B}_t := \int_0^t R_s^{-1}\sum_{j=1}^k B_s^{(j)}dB_s^{(j)}$, $t > 0$, is a continuous martingale with quadratic variation $\int_0^t |\mathbf{f}(s)|^d s = t$, where $\mathbf{f}$ is the vector-valued function with components $f_s^{(j)} = B_s^{(j)}/R_s$. Thus, by Lévy's characterization, Theorem 9.4, $\tilde{B}$ is a one-dimensional standard Brownian motion. Thus, letting $X_t := R_t^2$, $t \geq 0$, one has

$$dX_t = kdt + \sqrt{|X_t|}d\tilde{B}_t, t > 0, \quad X_0 = x,$$

for $x \geq 0$, is referred to as the *squared Bessel process starting at* x.

Example 8 (Diffusions on a Torus) On $\mathbb{R}^k$ $(k \geq 1)$ let $\boldsymbol{\mu}(\cdot)$, $\boldsymbol{\sigma}(\cdot)$ be Lipschitzian and *periodic*,

$$\boldsymbol{\mu}(\mathbf{x} + \boldsymbol{\nu}) = \boldsymbol{\mu}(\mathbf{x}), \quad \boldsymbol{\sigma}(\mathbf{x} + \boldsymbol{\nu}) = \boldsymbol{\sigma}(\mathbf{x}) \quad \forall\, \mathbf{x} \in \mathbb{R}^k,\ \forall\ \boldsymbol{\nu} \in \mathbb{Z}^k. \tag{13.35}$$

Consider the diffusion $\{\mathbf{X}_t : t \geq 0\}$, having these coefficients, so that

$$d\mathbf{X}_t = \boldsymbol{\mu}(\mathbf{X}_t)dt + \boldsymbol{\sigma}(\mathbf{X}_t)d\mathbf{B}_t.$$

Check that for any given $\boldsymbol{\nu} \in \mathbb{Z}^k$, $\mathbf{Y}_t := \mathbf{X}_t + \boldsymbol{\nu}$ satisfies the same equation

$$d\mathbf{Y}_t = \boldsymbol{\mu}(\mathbf{Y}_t)dt + \boldsymbol{\sigma}(\mathbf{Y}_t)d\mathbf{B}_t.$$

Therefore, Theorem 13.3 or Corollary 13.4 applies with $\mathcal{G}$ as the group of all translations by elements $\boldsymbol{\nu}$ of $\mathbb{Z}^k$. A maximal invariant is given by $\varphi(\mathbf{x}) := \mathbf{x} \mod \mathbf{1} \equiv (x^{(1)} \mod 1, \ldots, x^{(k)} \mod 1)$, $\mathbf{x} \in \mathbb{R}^k$. Thus, φ takes values in the torus $\mathcal{T} := \{\mathbf{x} \mod \mathbf{1} : \mathbf{x} \in \mathbb{R}^k\}$, which as a measurable space may be labeled as $([0, 1)^k, \mathcal{B}([0, 1)^k))$ ($\boldsymbol{\nu} \mod \mathbf{1} := 0$, $\boldsymbol{\nu} \in \mathbb{Z}^k$). In particular, $\mathcal{T}$ is compact. The Markov process $\mathbf{Z}_t := \varphi(\mathbf{X}_t)$, $t \geq 0$, is called *a diffusion on the* (unit) *torus.*

Example 9 (Brownian Motion on a Circle and on a Torus) For each $\mathbf{a} = (a^{(1)}, a^{(2)}, \ldots, a^{(k)})$ with $a^{(i)} > 0 (1 \leq i \leq k)$, one may define a diffusion $\{\mathbf{Z}_t, t \geq 0\}$ on the torus $\mathcal{T}_{\mathbf{a}} := \{\mathbf{x} \mod \mathbf{a} \equiv (x^{(1)} \mod a^{(1)}, \ldots, x^{(k)} \mod a^{(k)}) : \mathbf{x} \in \mathbb{R}^k\}$ by setting $\mathbf{Z}_t := \mathbf{X}_t \mod \mathbf{a}$. The transition probability density of $\{\mathbf{Z}_t, t \geq 0\}$ with respect to Lebesgue measure or $[0, a_1) \times \cdots \times [0, a_k)$ is easily shown to be

$$q(t; \mathbf{z}, \mathbf{z}') = \sum_{\boldsymbol{\nu} \in \mathbb{Z}^k} p(t; \mathbf{z}, \mathbf{z}' + (\nu^{(1)} a_1, \ldots, \nu^{(k)} a_k)), \quad (\mathbf{z}, \mathbf{z}' \in \mathcal{T}_{\mathbf{a}}, t > 0).$$

$$p(t; \mathbf{x}, \mathbf{y}) := (\frac{1}{\sqrt{2\pi t}})^k \exp\{-\frac{|\mathbf{y} - \mathbf{x}|^2}{2t}\} \qquad (\mathbf{x}, \mathbf{y} \in \mathbb{R}^k). \tag{13.36}$$

As a special case, the *Brownian motion on the unit circle* S^1 may be defined as $Z_t := B_t \mod 2\pi$, $t \geq 0$. Its transition probability density with respect to Lebesgue measure on S^1 (labeled as $[0, 2\pi)$) is

$$q(t; z, z') := \sum_{\nu \in \mathbb{Z}} (2\pi t)^{-\frac{1}{2}} \exp\{-\frac{1}{2t}(z' + 2\pi\nu - z)^2\}, \quad z, z' \in [0, 2\pi). \tag{13.37}$$

13.3 Reflecting Diffusions

Let $\mu(\cdot)$ and $\sigma(\cdot)$ be locally Lipschitzian on $\mathbb{R}$, such that the diffusion with these coefficients is nonexplosive. First assume also that $\mu(\cdot)$ is an *odd* function and $\sigma(\cdot)$ is an *even* function,

$$\mu(-x) = -\mu(x), \quad \sigma(-x) = \sigma(x), \qquad (x \in \mathbb{R}). \tag{13.38}$$

Then a diffusion $\mathbf{X} := \{X_t : t \geq 0\}$, with these coefficients and $\mathbf{Y} := \{Y_t := -X_t : t \geq 0\}$, satisfy the same Itô equation

$$\begin{aligned} dX_t &= \mu(X_t)dt + \sigma(X_t)dB_t, \\ dY_t &= \mu(Y_t)dt + \sigma(Y_t)d\widetilde{B}_t, \end{aligned} \tag{13.39}$$

where $\{\widetilde{B}_t = -B_t : t \geq 0\}$, is also a standard Brownian motion, just as $\{B_t : t \geq 0\}$. Therefore, $\{X_t : t \geq 0\}$, and $\{Y_t : t \geq 0\}$, have the same transition probability $p(t; x, dy)$. This is equivalent to saying that

$$p(t; x, B) = p(t; -x, -B), \qquad B \in \mathcal{B}(\mathbb{R}). \tag{13.40}$$

In other words, the transition probability is invariant (i.e., satisfies (13.28)) under the reflection group $\{g, e\}$, where $gx = -x$. A maximal invariant under this group is $|x|$ (another, equivalent, one is x^2). Hence, by Theorem 13.3, $Z_t := |X_t|, t \geq 0$, is a Markov process on $S = [0, \infty)$. It is called a diffusion on $[0, \infty)$ having coefficients $\mu(\cdot), \sigma(\cdot)$, with *reflection* at 0. If the process $\mathbf{X}$ has a transition probability density[4] $p(t; x, y)$ then Z has the transition probability density

$$\begin{aligned} q(t; x, y) &:= p(t; x, y) + p(t; x, -y) \\ &= p(t; x, y) + p(t; -x, y) \qquad (x > 0, y > 0, t > 0) \end{aligned} \tag{13.41}$$

satisfying the backward equations $\frac{\partial q}{\partial t} = Aq$.

Next, suppose one is given $\mu(\cdot)$ and $\sigma(\cdot)$ on $[0, \infty)$, which are locally Lipschitz, $\sigma^2(\cdot) > 0$. Extend them to $\mathbb{R}$ by (13.38). Change $\mu(0)$ to 0. As pointed out in Remark 13.1, the diffusion is well defined as a Feller diffusion (see Chapter 21), and (13.40), (13.41) hold.

Remark 13.3 It follows that $q(t; x, y)$ satisfies the same equation in (t, x) as $p(t; x, y)$. Namely,

$$\frac{\partial q(t; x, y)}{\partial t} = Aq(t; x, y), \qquad 0 < x, y < \infty \ (t > 0), \tag{13.42}$$

and, by (13.41), satisfies the *boundary condition*

$$\lim_{x \downarrow 0} \frac{\partial q}{\partial x}(t; x, y) = 0 \qquad (t > 0, y > 0). \tag{13.43}$$

Example 10 (Reflecting Brownian Motion on $[0, \infty)$*)* If B_t^x, $t \geq 0$, is a standard Brownian motion starting at $x > 0$, then $Z_t^x := |B_t^x|, t \geq 0$, is a *reflecting Brownian motion on* $[0, \infty)$ whose transition probability density is

$$q(t; x, y) := p(t; x, y) + p(t; x, -y), \qquad (x > 0, y > 0, t > 0)$$

[4] See Friedman (1964) for existence of a transition probability density.

$$p(t; x, y) := (2\pi t)^{-\frac{1}{2}} \exp\{-\frac{1}{2t}(y - x)^2\}. \tag{13.44}$$

Note that

$$\begin{aligned}
\left.\frac{\partial q(t; x, y)}{\partial x}\right|_{x\downarrow 0} &= 0 \qquad\qquad (t > 0, y > 0), \\
\left.\frac{\partial q(t; x, y)}{\partial x}\right|_{y\downarrow 0} &= 0 \qquad\qquad (t > 0, x > 0).
\end{aligned} \tag{13.45}$$

The relations (13.45) also hold for the general nonsingular reflecting diffusion on $[0, \infty)$. In the case of Brownian motion with drift $\mu \neq 0$, one may check that the process $Y_t = |\mu t + B_t|, t \geq 0$, does not have the Markov property (Exercise 1).

Remark 13.4 Suppose $\mu(x), \sigma(x) \neq 0$ are bounded, locally Lipschitzian, specified on $[0, \infty)$. If $\mu(0) \neq 0$, one may still define $\sigma(x), \mu(x)$ on $\mathbb{R}$ satisfying (13.38) for $x \neq 0$. In this case, $\mu(\cdot)$ is not continuous at 0. However, as noted in Remark 13.1, there still exists a symmetric diffusion with these extended coefficients (one may, in fact, modify $\mu(\cdot)$ to let $\mu(0) = 0$, without changing Feller's generator $A = \frac{d}{dm(x)} \frac{d}{ds(x)}$). Thus, a reflecting Brownian motion on $[0, \infty)$ may be constructed again as $Z_t := |X_t|, t \geq 0$. Alternatively, one may use pde to obtain the fundamental solution $q(t; x, y)$ satisfying (13.42), (13.43).

Example 11 (Reflecting Diffusions on $[0, 1]$*)* Suppose that Lipschitzian coefficients $\mu(\cdot)$ and $\sigma(\cdot) > 0$ are given on $[0, 1]$. Assume first that

$$\mu(0) = 0, \quad \mu(1) = 0. \tag{13.46}$$

Extend these coefficients to $[-1, 1]$ by setting

$$\mu(-x) := -\mu(x), \quad \sigma(-x) := \sigma(x), \qquad 0 < x \leq 1. \tag{13.47}$$

Now extend $\mu(\cdot), \sigma(\cdot)$ to all $\mathbb{R}$ *periodically with period 2:*

$$\mu(x + 2\nu) = \mu(x), \quad \sigma(x + 2\nu) = \sigma(x), \qquad -\infty < x < \infty, \ \nu \in \mathbb{Z}. \tag{13.48}$$

Let $X_t, t \geq 0$, be a diffusion on $\mathbb{R}$ with these extended coefficients. Then $Z_t := X_t \bmod 2$, $t \geq 0$, is a diffusion on the circle $\{x \bmod 2 : x \in \mathbb{R}\}$. Hence, $Z_t - 1$ is a Markov process on $[-1, 1)$, which is symmetric, i.e., $Z_t - 1$ and $-Z_t + 1$ have the same distribution (when -1 and $+1$ are identified). Thus, $Y_t := |Z_t - 1|$ is a Markov process on $[0, 1]$, which we call the *reflecting diffusion on* $[0, 1]$ with coefficients $\mu(\cdot), \sigma(\cdot)$, satisfying (13.47). One may also express $Y_t = \varphi(X_t)$, where φ is a maximal invariant under the group $\mathcal{G}$ generated by $\theta_0 =$ *reflection around* 0, $\theta_0 x = -x$, and $g_1 =$ *translation by* 2. The group $\mathcal{G}$ is also generated by θ_0 and the *reflection* θ_1 *around* $1 : \theta_1 x = 2 - x \ (= 1 + 1 - x)$. Identified as the orbit of x,

$\varphi(x) \approx o(x) = \{\pm x + 2\nu : \nu \in \mathbb{Z}\}$ $(x \in \mathbb{R})$, since the map $x \to o(x)$ on [0, 1] is one-to-one and onto the space of orbits, the reflecting diffusion on [0, 1] may also be defined in this manner. In particular, the transition probability density $q(t; x, y)$ of Y_t, $t \geq 0$, is given in terms of the transition probability density $p(t; x, y)$ of the unrestricted diffusion as (Exercise 10)

$$q(t; x, y) = \sum_{\nu \in \mathbb{Z}} p(t; x+1, y+1+2\nu) + \sum_{\nu \in \mathbb{Z}} p(t; x+1, -y+1+2\nu)\ x, y \in [0, 1]), t > 0. \tag{13.49}$$

As pointed out in Remark 13.4, given any bounded Lipschitzian $\mu(\cdot), \sigma(\cdot) > 0$ on [0, 1], one can simply change $\mu(\cdot)$, if necessary, so that $\mu(0) = \mu(1) = 0$, and use the above construction for a symmetric diffusion with $\mu(\cdot)$ having possible discontinuities at integer points.

Example 12 (Reflecting Brownian Motion on $[0, 1]$*)* For a one- dimensional standard Brownian motion B_t, $t \geq 0$, the reflecting diffusion $Y_t = |Z_t - 1|$, where $Z_t = B_t \mod 2$, has the transition probability density (13.49) with $p(t; x, y) = (2\pi t)^{-\frac{1}{2}} \exp\{-(x-y)^2/2t\}$.

Example 13 (Reflecting Brownian Motion on $[0, 1]^k, k \geq 2$*)* Given a standard k-dimensional Brownian motion $\mathbf{B}_t$, $t \geq 0$, the reflecting Brownian motion on $[0, 1]^k$ is $\mathbf{Y}_t = (Y_t^{(1)}, \ldots, Y_t^{(k)})$, $t \geq 0$, where $Y_t^{(i)} = |Z_t^{(i)} - 1|$, $Z_t^{(i)} := B_t^{(i)} \mod 2$.

Example 14 (General Multidimensional Diffusions with Reflection on a Half Space) Let $\boldsymbol{\mu}(\mathbf{x})$, $\boldsymbol{\sigma}(\mathbf{x})$ be prescribed on $\overline{H}_k = H_k \cup \partial H_k$, where $H_k = \{\mathbf{x} \in \mathbb{R}^k, x^{(1)} > 0\}$, $\partial H_k = \{\mathbf{x} \in \mathbb{R}^k, x^{(1)} = 0\}$. Assume $\boldsymbol{\mu}(\cdot)$, $\boldsymbol{\sigma}(\cdot)$ to be Lipschitzian, $\boldsymbol{\sigma}(\cdot)$ symmetric, satisfying

$$\mu^{(1)}(\mathbf{x}) = 0, \quad \sigma_{1j}(\mathbf{x}) = 0, \quad 2 \leq j \leq k, \quad \text{for } \mathbf{x} \in \partial H_k. \tag{13.50}$$

Extend the coefficients to all of $\mathbb{R}^k$ by setting, for all $\mathbf{x} = (x^{(1)}, \ldots, x^{(k)})$, the vector $\overline{\mathbf{x}} = (-x^{(1)}, x^{(2)}, \ldots, x^{(k)})$, and letting

$$\begin{aligned}
\mu^{(1)}(\overline{\mathbf{x}}) &= -\mu^{(1)}(\mathbf{x}), \quad \sigma_{1j}(\overline{\mathbf{x}}) = -\sigma_{1j}(\mathbf{x}) \quad \text{for } 2 \leq j \leq k,\\
\mu^{(i)}(\overline{\mathbf{x}}) &= \mu^{(i)}(\mathbf{x}) \quad \text{for } 2 \leq i \leq k,\\
\sigma_{11}(x) = \sigma_{11}(\overline{\mathbf{x}}), &\quad \sigma_{ij}(\overline{\mathbf{x}}) = \sigma_{ij}(\mathbf{x}) \quad \text{for } 2 \leq i, j \leq k, \quad (\mathbf{x} \in \mathbb{R}^k).
\end{aligned} \tag{13.51}$$

Then writing $\overline{\mathbf{X}}_t = (-X_t^{(1)}, X_t^{(2)}, \ldots, X_t^{(k)})$ one has

$$\begin{aligned}
d\mathbf{X}_t &= \boldsymbol{\mu}(\mathbf{X}_t)dt + \boldsymbol{\sigma}(\mathbf{X}_t)d\mathbf{B}_t,\\
d\overline{X}_t^{(1)} &\equiv -dX_t^{(1)} = -\mu^{(1)}(\mathbf{X}_t)dt - \sum_{j=1}^{k} \sigma_{1j}(\mathbf{X}_t)dB_t^{(j)}
\end{aligned}$$

$$= \mu^{(1)}(\overline{\mathbf{X}}_t)dt + \sigma_{11}(\overline{\mathbf{X}}_t)d\overline{B}_t^{(1)} + \sum_{j=2}^{k} \sigma_{1j}(\overline{\mathbf{X}}_t)dB_t^{(j)}$$

$$= \mu^{(1)}(\overline{\mathbf{X}}_t)dt + \sigma_{11}(\overline{\mathbf{X}}_t)d\overline{B}_t^{(1)} + \sum_{j=2}^{k} \sigma_{1j}(\overline{\mathbf{X}}_t)dB_t^{(j)},$$

where $\overline{B}_t^{(1)} := -B_t^{(1)}$. Also, for $i \geq 2$,

$$d\overline{X}_t^{(i)} = dX_t^{(i)} = \mu^{(i)}(\mathbf{X}_t)dt + \sum_{j=1}^{k} \sigma_{ij}(\mathbf{X}_t)dB_t^{(j)}$$

$$= \mu^{(i)}(\overline{\mathbf{X}}_t)dt + \sigma_{i1}(\overline{\mathbf{X}}_t)d\overline{B}_t^{(1)} + \sum_{j=2}^{k} \sigma_{ij}(\overline{\mathbf{X}}_t)dB_t^{(j)}.$$

In other words, $\overline{\mathbf{X}}_t$ satisfies the Itô equation $d\overline{\mathbf{X}}_t = \boldsymbol{\mu}(\overline{\mathbf{X}}_t)dt + \boldsymbol{\sigma}(\overline{\mathbf{X}}_t)d\overline{\mathbf{B}}_t$, where $\overline{\mathbf{B}}_t = (-B_t^{(1)}, B_t^{(2)}, \ldots, B_t^{(k)})$ is a standard Brownian motion on $\mathbb{R}^k$. Hence, $\mathbf{X}_t$ and $\overline{\mathbf{X}}_t$ have the same transition probability. It follows that $\mathbf{Y}_t := (|X_t^{(1)}|, X_t^{(2)}, \ldots, X_t^{(k)})$, $t \geq 0$, is a Markov process on $\overline{H}_k$. We will call $\mathbf{Y}_t$, $t \geq 0$, the *reflecting diffusion on $\overline{H}_k$ with reflection on ∂H_k.*

Remark 13.5 It is known[5] that the unrestricted diffusion, whose coefficients satisfy (13.50), (13.51), has a transition probability density $p(t; \mathbf{x}, \mathbf{y})$ if, in addition to the assumed smoothness of the coefficients, $\sigma(\cdot)$ is assumed nonsingular. Then the transition probability $q(t; \mathbf{x}, \mathbf{y})$ of the reflecting diffusion $\mathbf{Y}_t$ is given by

$$\begin{aligned} q(t; \mathbf{x}, \mathbf{y}) &= p(t; \mathbf{x}, \mathbf{y}) + p(t; \mathbf{x}, \overline{\mathbf{y}}) \\ &= p(t; \mathbf{x}, \mathbf{y}) + p(t; \overline{\mathbf{x}}, \mathbf{y}), \qquad (\mathbf{x}, \mathbf{y} \in H_k). \end{aligned} \tag{13.52}$$

This implies that q satisfies $\partial q/\partial t = Aq$ and the *Neumann boundary condition*

$$\left.\frac{\partial q(t; x, y)}{\partial x^{(1)}}\right|_{x^{(1)}=0} = 0 \quad (\mathbf{y} \in H_k), \qquad \left.\frac{\partial q(t; \mathbf{x}, \mathbf{y})}{\partial y^{(1)}}\right|_{y^{(1)}=0} = 0, \quad (\mathbf{x} \in H_k). \tag{13.53}$$

As indicated in earlier remarks (see Remark 13.4), the condition '$\mu^{(1)}(\mathbf{x}) = 0$ for $\mathbf{x} \in \partial H$', may be dispensed with. To understand the role of the second condition '$\sigma_{1j}(\mathbf{x}) = 0$ for $2 \leq j \leq k$ for $\mathbf{x} \in \partial H_k$', note first that in view of the assumed symmetry of the matrix $\boldsymbol{\sigma}(\cdot)$, this condition is equivalent to

[5] See Friedman (1964).

$$a_{1j}(\mathbf{x}) = 0 \quad \text{for} \quad 2 \le j \le k, \ \mathbf{x} \in \partial H_k, \tag{13.54}$$
$$\mathbf{a}(\mathbf{x}) := \boldsymbol{\sigma}(\mathbf{x})\boldsymbol{\sigma}'(\mathbf{x}).$$

The following example shows that (13.54) ensures that the reflection is in the "normal" direction as opposed to a *"conormal" direction* as explained below.

We have proved the following result for the process $\{\mathbf{Y}_t : t \ge 0\}$.

Theorem 13.5 *Given locally Lipschitz coefficients $\boldsymbol{\mu}(\cdot)$, $\boldsymbol{\sigma}(\cdot)$, $\boldsymbol{\sigma}(\cdot)$ symmetric and positive definite, on the half space $\overline{H}_k = \{\mathbf{x} = (x^{(1)}, x^{(2)}, \dots, x^{(k)}) \in \mathbb{R}^k : x^{(1)} \ge 0\}$, satisfying (13.17), there exists a diffusion $\{\mathbf{Y}(t) : t \ge 0\}$, on the half-space $(\mathbf{H}_k)$ reflecting at the boundary $\partial \overline{H}_k = \{\mathbf{x} \in \mathbb{R}^k : x^{(1)} = 0\}$, whose transition probability density $q(t; \mathbf{x}, \mathbf{y})$ obeys the backward equation $\frac{\partial q}{\partial t} = Aq$ in H_k, where for $\mathbf{a}(\cdot) = \sigma(\cdot)\boldsymbol{\sigma}(\cdot)'$,*

$$Aq = \frac{1}{2} \sum_{1 \le i,j \le k} a_{ij}(\mathbf{x}) \frac{\partial^2 q(t; \mathbf{x}, \mathbf{y})}{\partial x^{(i)} x^{(j)}} + \sum_{1 \le i \le k} \mu^{(i)}(\mathbf{x}) \frac{\partial q(t; \mathbf{x}, \mathbf{y})}{\partial x^{(i)}}. \tag{13.55}$$

and the Neumann boundary condition (13.53) on $\partial \overline{H}_k$.

Example 15 (Diffusion on a Half-Space of $\mathbb{R}^k$ with Conormal Reflection at the Boundary) Consider an arbitrary half-space

$$\overline{H}_{k,\gamma} = \{\mathbf{x} = (x^{(1)}, x^{(2)}, \dots, x^{(k)}) \in \mathbb{R}^k, \gamma \cdot \mathbf{x} \ge 0\},$$

where γ is a unit vector in $\mathbb{R}^k$ and $\gamma \cdot \mathbf{x} = \sum_{1 \le i \le k} \gamma^{(i)} x^{(j)}$ is the usual inner product in $\mathbb{R}^k$. Suppose we are given a locally Lipschitz drift $\boldsymbol{\nu}(\cdot) \in \overline{H}_{k,\gamma}$, and a constant symmetric positive definite matrix $\mathbf{a} = ((a_{ij}))$. Let O be an orthogonal transformation that maps $\mathbf{a}^{\frac{1}{2}}\gamma/|\mathbf{a}^{\frac{1}{2}}\gamma|$ to $\mathbf{e} = (1, 0, 0, \dots, 0)$. Let $\{\mathbf{Y}_t : t \ge 0\}$, be the diffusion on $\overline{H}_k$ given by Theorem 13.5 with drift $\boldsymbol{\mu}(\cdot) = (\mathbf{a}^{\frac{1}{2}} O)^{-1} \nu(\cdot)$, and the diffusion matrix $\mathbf{I}$ (identity matrix). It is simple to check that

$$\mathbf{X}(t) = \mathbf{a}^{\frac{1}{2}} O Y_t, \quad t \ge 0,$$

is a diffusion on $\overline{H}_{k,\gamma}$ with drift $\boldsymbol{\nu}$ and diffusion matrix $\mathbf{a}$, $\mathbf{a}^{\frac{1}{2}} O$ being a diffeomorphism on $\overline{H}_k$ onto $\overline{H}_{k,\gamma}$. Expressed in terms of this transformation, the backward equation for the transition probability density p, say, of the process $\{\mathbf{X}_t : t \ge 0\}$ is $\frac{\partial p}{\partial t} = Ap$, where

$$Ap = \frac{1}{2} \sum_{1 \le i,j \le k} a_{ij} \frac{\partial^2 p}{\partial x^{(i)} \partial x^{(j)}} + \sum_{1 \le i \le k} \mu^{(i)} \frac{\partial p}{\partial x^{(i)}}, \tag{13.56}$$

and the boundary condition becomes (Exercise 12)

$$(\mathbf{a}\gamma)\cdot(\mathrm{grad})_x p(t;x,y)=0, \text{ or } \sum_{1\le i\le k}\Big(\sum_{1\le j\le k} a_{ij}\gamma_i\Big)\frac{\partial p(t;\mathbf{x},\mathbf{y})}{\partial x^{(i)}}=0. \tag{13.57}$$

The boundary condition (13.57) is called a *conormal boundary condition* (see Exercise 12).

Remark 13.6 Suppose one is given an open set $G \subset \mathbb{R}^k$ ($k > 1$) with (1) a smooth boundary ∂G, (2) an infinitesimal generator $\mathbf{A}$ with coefficients $\boldsymbol{\mu}(\cdot)$ and $\boldsymbol{\sigma}(\cdot)$ Lipschitzian on $\overline{G}$, $\boldsymbol{\sigma}(\cdot)$ having eigenvalues bounded away from zero and infinity, and (3) a continuous field of unit vectors $\boldsymbol{\gamma}(\mathbf{x})$ ($\mathbf{x} \in \partial G$) such that $\boldsymbol{\gamma}(\mathbf{x})\cdot\mathbf{n}(\mathbf{x}) \ge C > 0$ for all $\mathbf{x} \in \partial G$, where $\mathbf{n}(\mathbf{x})$ is the unit outward normal to ∂G at $\mathbf{x}$. Then there exists a reflecting diffusion on $\overline{G}$ whose transition probability density $q(t;\mathbf{x},\mathbf{y})$ satisfies[6]

$$\begin{aligned}
&\frac{\partial q(t;\mathbf{x},\mathbf{y})}{\partial t}=\mathbf{A}q(t;\mathbf{x},\mathbf{y}) \qquad (\mathbf{x},\mathbf{y}\in G, t>0),\\
&\boldsymbol{\gamma}(\mathbf{x})\cdot\mathbf{grad}_{\mathbf{x}}\, q(t;\mathbf{x},\mathbf{y})=0 \qquad (\mathbf{x}\in\partial G, \mathbf{y}\in G, t>0).
\end{aligned}$$

Exercises

For the exercises below, inaccessible boundaries are those that, starting from the interior, cannot be reached in finite time.

1. For $\mu \neq 0$, (a) show that $Y_t = |\mu t + B_t|, t \ge 0$, does not have the Markov property. [*Hint*:Compute the transition probabilities $p(t, y_0, y_1)$ for Y on $[0,\infty)$ and show that the Chapman–Kolmogorov equations fail for $p(t+t, , 0, 0)$.] (b) Using the method of this chapter, how would one construct a diffusion on $[0,\infty)$ with drift μ, diffusion coefficient 1, and a reflecting boundary condition $\frac{\partial p(t;x,y)}{\partial x}|_{x=0}=0$. Prove that the diffusion is transient.
2. Let $\mu(\cdot), \sigma(\cdot) > 0$ be locally Lipschitzian on (a,b), $a<b$, with either $a>-\infty$ or $b<\infty$, or both a and b are finite. Define the scale and speed functions $s(x)$ and $m(x)$ as in (13.4). Assume (13.3) holds.
   (i) Prove that a and b are inaccessible.
   (ii) Prove that the diffusion is recurrent if and only if $s(a) := \lim_{x\downarrow a} s(x) = -\infty$ and $s(b) := \lim_{x\uparrow b} s(x) = \infty$.
   (iii) Prove that the diffusion is positive recurrent if and only if $m(a) > -\infty$ and $m(b) < \infty$. [*Hint*: Use $A \equiv \frac{1}{2}\sigma^2(x)(d^2/dx^2) + \mu(x)(d/dx) =$

[6] See Ikeda and Watanabe (1989), pp. 217-232, or Stroock and Varadhan (1979), pp. 147-225. For the special case of Brownian motion on the two-dimensional *closed unit disc* $\overline{G}$ with general oblique (or, non-conormal) reflection, see McKean (1969), pp. 126-132.

$(d/dm(x))(d/ds(x))$, compute (1) $P(X^x$ reaches c before $d)$ as $[s(x) - s(c)]/[s(d) - s(c)]$, $a < c < x < d < b$, and (2) $\gamma(x) :=$ expected time for X^x to reach $\{c, d\}$, as the solution of $A\gamma(x) = -1$, $\gamma(c) = \gamma(d) = 0$.]

(iv) Derive (13.3) from Feller's criterion for explosion on $(-\infty, \infty)$ (Theorem 12.2), using a diffeomorphism $\varphi : (a, b) \to \mathbb{R}$.

3. Determine values of the parameter σ under which the zero-seeking Brownian motion given in Example 1 is transient, null-recurrent, and ergodic.
4. Consider the diffusion on (0, 1) with coefficients $\mu(x) = C_1(1 - x) - C_2 x$, $\sigma(x) = (x(1 - x))^{\frac{1}{2}}$, $C_1 \geq 0, C_2 \geq 0$.

   (i) Show that 0 and 1 are inaccessible if and only if $C_1 \geq \frac{1}{2}, C_2 \geq \frac{1}{2}$.
   (ii) If $C_1 \geq \frac{1}{2}, C_2 \geq \frac{1}{2}$, show that the diffusion on (0, 1) is positive recurrent.
   (iii) Check that $m(0)$ and $m(1)$ are finite if $c_1 > 0, c_2 > 0$. Hence, with the additional condition of recurrence, it is not enough for positive recurrence that the speed measure is finite at the boundary.

5. Prove that in Example 3, the distribution $\overline{p}(t; \mathbf{x}, d\mathbf{y})$ on the boundary $\{\mathbf{y} \in \mathbb{R}^k, y^{(1)} = 0\}$, with $\mathbf{x}$ away from the boundary (i.e., $x^{(1)} > 0$), as given in (13.27), has for $k = 2$ then density (with respect to Lebesgue measure $dy^{(2)}$ on the boundary) $\overline{p}(t; \mathbf{x}, (0, y^{(2)})) = \frac{x^{(1)}}{\pi}[(x^{(1)})^2 + (y^{(2)} - x^{(2)})^2]^{-1} \exp\{-[(x^{(1)})^2 + (y^{(2)} - x^{(2)})^2]/2t\}$ which converges, as $t \uparrow \infty$, to the Cauchy density $f(y^{(2)}; x^{(1)}, x^{(2)}) = (\pi x^{(1)})^{-1}[1 + (\frac{y^{(2)}}{x^{(1)}} - \frac{x^{(2)}}{x^{(1)}})^2]^{-1}$, $y^{(2)} \in \mathbb{R}$ $(x^{(1)} > 0, x^{(2)} \in \mathbb{R})$.
6. (i) Find conditions on $\alpha(\cdot) > 0$ and $\beta(\cdot)$ such that the diffusion on $(0, \infty)$ with generator (13.33) has inaccessible boundaries.
   (ii) In the case of inaccessible boundaries, find conditions on $\alpha(\cdot)$, $\beta(\cdot)$ such that the diffusion is (a) null recurrent, or (b) positive recurrent, or (c) transient.
7. Consider Example 6.

   (i) Show that $R = \{R_t : t \geq 0\}$ is a Markov process on $S = (0, \infty)$ and that for twice continuously differentiable functions f with compact support, one has $\frac{d}{dt} T_t^R f = A^R T_t^R f$, where T_t^R, $t \geq 0$, are the transition operators for $\{R_t : t \geq 0\}$. ("0" is an example of an *inaccessible boundary* for the process R).
   (ii) Since the range of $|B_t|$ is $[0, \infty)$, one may consider the radial Brownian motion on $\overline{S} = [0, \infty)$. Show that $R = \{R_t : t \geq 0\}$ is a Markov process on the state space $\overline{S} = [0, \infty)$, with the same transition probability $p(t; r, dr)$ for $r > 0$ as in (i), but with $p(t; 0, (0, \infty)) = 1$ for all $t > 0$ for $k \geq 2$. (In Feller's boundary classification, "0" is an example of an *entrance boundary* for the process R).

8. (i) If $a < b$ are finite, and $\mu(x) = 0, \sigma^2(x) > 0$ on (a, b), then show that the boundaries a, b are accessible. (ii) If $a = -\infty, b = +\infty, \mu(x) = 0, \sigma^2(x) > 0$ on $(-\infty, \infty)$, then show that the boundaries are inaccessible and, indeed, the

diffusion is recurrent. (iii) Under the additional condition on $\sigma^2(\cdot)$ in (ii), is the diffusion recurrent?

9. (*Diffusion on the Unit Sphere*) Suppose a diffusion on $\mathbb{R}^k$, $k > 1$, is nonexplosive with locally Lipschitzian coefficients satisfying $\boldsymbol{\mu}(c\mathbf{x}) = c\boldsymbol{\mu}(\mathbf{x})$, $\boldsymbol{\sigma}(c\mathbf{x}) = \pm c\boldsymbol{\sigma}(\mathbf{x})\ \forall\, c \in \mathbb{R}$.

   (i) Show that its transition probability is invariant under the group $\mathcal{G}_+$ of transformations $g_c(\mathbf{x}) = c\mathbf{x}$, $c > 0$, and that a maximal invariant under this group is $\varphi(\mathbf{x}) = \mathbf{x}/|\mathbf{x}|$ (if $\mathbf{x} \neq \mathbf{0}$), $\varphi(\mathbf{0}) = \mathbf{0}$.
   (ii) Consider the example $\boldsymbol{\mu}(\mathbf{x}) = a\mathbf{x}$, and for some $\gamma \neq 0$, $\sigma_{i,i}(\mathbf{x}) = \gamma x^{(i+1)}$, $1 \le i \le k-1$, $\sigma_{kk}(\mathbf{x}) = -x^{(1)}$, $\sigma_{i,i+1}(\mathbf{x}) = -\gamma x^{(i)}$, $1 \le i \le k-1$, $\sigma_{k1} = x^{(k)}$, and all other coefficients are zero. Show that $\boldsymbol{\sigma}(\mathbf{x})\mathbf{x} = \mathbf{0}$ for all $\mathbf{x}$, and that $R_t^2 := |\mathbf{X}_t^{\mathbf{x}}|^2$ equals $|\mathbf{x}|^2 \exp\{(2\gamma^2 - a^2)t\}$, $t \ge 0$, a deterministic process. Show also that if $2\gamma^2 = a^2$, then, for every initial $\mathbf{x} \neq 0$, $\mathbf{X}_t^{\mathbf{x}}$, $t \ge 0$, is a Markov process on the sphere $\{\mathbf{y} \in \mathbb{R}^k : |\mathbf{y}| = |\mathbf{x}|\}$.

10. Verify (13.49) and establish the backward and forward boundary conditions: $\frac{\partial q}{\partial x} = 0$, $x = 0, 1$, and $\frac{\partial q}{\partial y} = 0$, $y = 0, 1$, respectively.
11. (a) Consider k-dimensional standard Brownian motion on the orthant $O_k = \{\mathbf{x} \in \mathbb{R}^k : x^{(i)} \ge 0, 1 \le i \le k\}$, with normal reflection at the boundary. That is, the process is $\mathbf{X}_t^{\mathbf{x}} = \{(|B_t^{(1)}|, \dots, |B_t^{(k)}|), \mathbf{x} = (x^{(1)}, \dots, x^{(k)}), t \ge 0$. Here, $\mathbf{B}_t = (B_t^{(1)}, \dots, B_t^{(k)})$, $t \ge 0$, is a k-dimensional standard Brownian motion. (i) Compute its transition probability density. (ii) Show that the process is null-recurrent for $k \ge 1$.
12. In Example 15, verify the assertion that the diffeomorphism $\mathbf{a}^{\frac{1}{2}}O$ on $\overline{H}_k$ onto $\overline{H}_{k,\gamma}$ transforms the diffusion $\{\mathbf{Y}_t : t \ge 0\}$ on $\overline{H}_k$ with drift $\mu(\cdot)$ and diffusion matrix $\mathbf{I}$, and normal reflection at the boundary $\{x : x^{(1)} = 0\}$, to a process $\{X_t : t \ge 0\}$ on $\overline{H}_{k,\gamma}$ with drift $\nu(\cdot) = \mathbf{a}^{\frac{1}{2}}O\mu(\cdot)$, and diffusion matrix $\mathbf{a}$, and the conormal boundary condition in (13.57).

Chapter 14
The Speed of Convergence to Equilibrium of Discrete Parameter Markov Processes and Diffusions

Coupling methods and discrete renewal theory are used to derive criteria for polynomial as well as exponential rates of convergence to equilibrium for Harris recurrent Markov processes in discrete time and diffusions in continuous time.

In this chapter, based mostly on Wasielak (2009) and Bhattacharya and Wasielak (2012), criteria for polynomial and exponential rates of convergence of Markov processes to equilibrium are derived both for discrete time Harris positive recurrent Markov processes, as well as d-dimensional diffusions, $d \geq 1$, using the method of coupling. Lindvall (2002) used this method, which he attributes to Jim Pitman, to obtain polynomial rates of convergence of one-dimensional diffusions to equilibrium. A more general description of his method is given in Lindvall (2002). Some criteria for polynomial convergence of multidimensional diffusions were derived by Davies (1986), also using estimates of moments of the coupling time. To state the necessary notions from renewal theory (see, e.g., Bhattacharya and Waymire (2021), Chapter 8), let $\{Y_n : n = 1, 2, \dots\}$ be a sequence of i.i.d. strictly positive integer-valued random variables, with lattice span one, having the probability mass function (pmf) $f(j)$, $j \geq 1$; and let Y_0 be another non-negative integer-valued random variable independent of $\{Y_n : n = 1, 2, \dots\}$, Y_0 is the *delay distribution*. This refers to the times Y_n needed to repair an item, with Y_0 being the delay in starting the repair. Assume $\mu = \mathbb{E}Y_1 < \infty$. Let

$$S_n := Y_0 + \cdots + Y_n, \quad N_n := \inf\{k \geq 0 : S_k \geq n\}$$

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_14

N_n is the nth *renewal epoch*, $S_{N_n}, n \geq 0$ is the *renewal process*, and $R_n = S_{N_n} - n$ is the *residual lifetime* (or overshoot) of the last renewal prior to the nth epoch. It helps to think about the renewal process as a point process $\{S_n\}$ on the integers with the points $S_n, n \geq 0$, laid out with blue dots on an axis. It is well known that $R_n, n \geq 1$, is a Markov process on the space $\mathbb{Z}_+$ of non-negative integers, with the unique invariant distribution given by the integrated tail distribution[1]

$$g(j) = \sum_{i>j} \frac{f(i)}{\mu}, \quad j = 0, 1, 2, \ldots.$$

This is to be coupled with renewals of the process

$$\tilde{S}_n = \tilde{Y}_0 + \tilde{Y}_1 + \cdots + \tilde{Y}_n$$

where $\tilde{Y}_n, n \geq 1$, are i.i.d. with the same common distribution as $Y_n, n \geq 1$, and $\tilde{Y}_0$ an independent delay. Let $\tilde{N}_n$, $\tilde{S}_{\tilde{N}_n}$, $\tilde{R}_n$, be defined for this sequence as for the original sequence; but $\tilde{Y}_0$ is assumed to have the invariant distribution g. One may think of this point process $\{\tilde{S}_n, n \geq 0\}$ placed with red dots on the same axis as $\{S_n : n \geq 0\}$. Coupling occurs at a time T when a red dot falls on a blue dot, or, equivalently,

$$T = \inf\{n \geq 0 : R_n = \tilde{R}_n\} \equiv \inf\{n \geq 0 : S_{N_n} = \tilde{S}_{\tilde{N}_n}\}. \tag{14.1}$$

For estimating moments of T the following lemma will be used (See Lindvall (2002), Lemma (4.1)).

Lemma 10 *(i) For each $\rho > 0$, there exists a constant $C = C(\rho)$ such that $\mathbb{E}R_n \leq C + \rho n$. (ii) Suppose $\mathbb{E}(Y_1^\beta) < \infty$ for some $\beta \geq 1$. Then*

$$\mathbb{E}R_n^\beta \leq n\mathbb{E}Y_1^\beta \textit{ for all } n \geq 1.$$

(iii) If φ is an increasing non-negative function on $\mathbb{Z}_+$, then for all $n \geq 1$,

$$\mathbb{E}\varphi(R_n) \leq \mathbb{E}\varphi(Y_1)n.$$

Proof (i). Since $R_n = S_{N_n} - n$, using Wald's second identity (see, e.g., Bhattacharya and Waymire (2021), Chapter 19 and p. 142, or BCPT,[2] p. 72), one has

$$\mathbb{E}R_n = \mu\mathbb{E}N_n - n.$$

[1] Bhattacharya and Waymire (2021), p. 108.

[2] Throughout, BCPT refers to Bhattacharya and Waymire (2016), A Basic Course in Probability Theory.

By the renewal theorem[3]

$$\mu \mathbb{E} N_n / n \to 1 \text{ as } n \to \infty,$$

i.e., $\mathbb{E} R_n / n \to 0$. Therefore, given any $\rho > 0$, however small, there exists n_1 such that

$$\mathbb{E} R_n < \rho_n \text{ for } n \geq n_1.$$

Take $C = \max\{\mathbb{E} R_n : n \leq n_1\}$.

(iii) One has,

$$\begin{aligned}
P(R_n = k) &= \sum_{j \geq 1} P(R_n = k, S_j < n, S_{j+1} \geq n) \\
&= \sum_{j \geq 1} \sum_{0 \leq i \leq n-1} P(S_j = i, Y_{j+1} = n + k - i) \\
&= \sum_{j \geq 1} \sum_{0 \leq i \leq n-1} f(n + k - i) P(S_j = i) \\
&= \sum_{0 \leq i \leq n-1} (\sum_{j \geq 1} P(S_j = i)) f(n + k - i) \\
&\leq \sum_{0 \leq i \leq n-1} f(n + k - i), \; k \geq 0.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\varphi(R_n) &= \sum_{k \geq 0} \varphi(k) \sum_{0 \leq i \leq n-1} f(n + k - i) \\
&= \sum_{0 \leq i \leq n-1} (\sum_{k \geq 0} \varphi(k) f(n + k - i)) \\
&\leq \sum_{0 \leq i \leq n-1} \sum_{k \geq 0} \varphi(n + k - i) f(n + k - i) \\
&= \sum_{0 \leq i \leq n-1} \sum_{k' \geq n-i} \varphi(k') f(k') \\
&\leq n \mathbb{E} \varphi(Y_1).
\end{aligned}$$

We have used the facts that (a)

[3] See, e.g., Bhattacharya and Waymire (2021), Chapter 8, Theorem 8.5.

$$\sum_{j\geq 1} P(S_j = i) = \mathbb{E}(\sum_{j\geq 1} \mathbf{1}_{[S_j=i]}) \leq 1 \text{ for all } i,$$

and (b)

$$\varphi(k) \leq \varphi(n+k-i) \text{ for } i < n.$$

Part (ii) is a special case of (iii) with $\varphi(k) = k^\beta$. ∎

The lemma is helpful in the proof of the following result of Lindvall (2002), Theorem (4.2), which, in turn, will be used to derive the main results of this chapter.

Theorem 14.1 (Lindvall) *Let $\{Y_n : n \geq 1\}$ be an i.i.d. sequence of strictly positive integer-valued random variables with lattice span 1, and Y_0 a non-negative integer-valued random variable independent of $\{Y_n : n \geq 1\}$. Assume $\mathbb{E}Y_0^\beta < \infty$, $\mathbb{E}Y_1^\beta < \infty$, for some $\beta \geq 1$. Let $\{\tilde{Y}_n : n \geq 1\}$, be a sequence independent of $\{Y_n : n \geq 1\}$, but with the same common distribution as Y_1, and with $\tilde{Y}_0$ having the invariant distribution of the residual renewal process $\tilde{R}_n$. Assume $\mathbb{E}\tilde{Y}_0^\beta < \infty$. Then $\mathbb{E}T^\beta < \infty$, where T is the coupling time defined by (14.1).*

Remark 14.1 Before proceeding with the rather long proof, it is important to note that the case $\beta = 1$ simply follows from the fact that the Markov chain $(R_n, \tilde{R}_n)$, $n \geq 1$, is positive recurrent. For it has the unique invariant probability (mass function) $g(\cdot) \times g(\cdot)$.

Proof Following Lindvall (2002), Chapter 2, but with some more details from Wasielak (2009), we define stopping times involving the sequences $S, \tilde{S}$, and corresponding values of these sequences for a possible match. Recalling that[4]

$$P(R_n = 0) \to g(0) > 0 \text{ as } n \to \infty,$$

one can find $0 < \gamma < g(0)$, and an integer n_0 such that for $n \geq n_0$,

$$P(R_n = 0) > \gamma, \text{ for all } n \geq n_0.$$

Now define, recursively

$$\begin{aligned} &\tau = \inf\{j : B_j = 0\}, A_0 = S_{n_0}, \nu_0 = \inf\{j : \tilde{S}_j \geq S_{n_0}\}, \\ &B_0 = \tilde{S}_{\nu_0} - S_{n_0}, \\ &A_1 = \tilde{S}_{\nu_0+n_0}, \ \nu_1 = \inf\{j > \nu_0 : S_j \geq A_1\}, \ B_1 = S_{\nu_1} - A_1, \end{aligned} \tag{14.2}$$

[4] See Bhattacharya and Waymire (2021), (8.17).

$$A_{2n+1} = \tilde{S}_{\nu_{2n}+n_0},\ n \geq 0,\ A_{2n} = S_{\nu_{2n-1}+n_0},\ (n \geq 1),$$
$$\nu_{2n+1} = \inf\{j > \nu_{2n} : S_j \geq A_{2n+1}\},\ B_{2n+1} = S_{\nu_{2n+1}} - A_{2n+1}, n \geq 1,$$
$$\nu_{2n} = \inf\{j > \nu_{2n-1} : \tilde{S}_j \geq A_{2n}\},\ B_{2n} = \tilde{S}_{\nu_{2n}} - A_{2n}\ n \geq 1.$$

Note that $B_j = 0$ implies a coupling at the jth step. This stopping time is attained in τ steps of lengths B_j, $j \geq 0$. Therefore, writing

$$\begin{aligned} U_{2n+1} &= S_{\nu_{2n+1}} - \tilde{S}_{\nu_{2n}} \geq B_{2n+1}, \\ U_{2n+2} &= \tilde{S}_{\nu_{2n+2}} - S_{\nu_{2n+1}} \geq B_{2n+2},\ (n \geq 0), \end{aligned} \tag{14.3}$$

one has

$$\begin{aligned} T &\leq \tilde{S}_{\nu_0} - S_{n_0} + \sum_{1\leq j\leq\tau} U_j = \tilde{S}_{\nu_0} - S_{n_0} + \sum_{1\leq j<\infty} U_j \mathbf{1}_{[\tau\geq j]} \\ &= \tilde{S}_{\nu_0} - S_{n_0} + \sum_{1\leq j<\infty} U_j \mathbf{1}_{[\tau> j-1]}. \end{aligned} \tag{14.4}$$

Assume now that the β-th moments of Y_0, $\tilde{Y}_0$ and Y_1 are finite, $\beta \geq 1$,

$$\mathbb{E}Y_0^\beta \equiv \sum_{0\leq i<\infty} i^\beta h(i) < \infty, \tag{14.5}$$

where h is the pmf of Y_0,

$$\mathbb{E}(\tilde{Y}_0)^\beta = \sum_{1\leq i<\infty} i^\beta g(i) < \infty, \tag{14.6}$$

$$\mathbb{E}(Y_1)^\beta = \sum_{1\leq i<\infty} i^\beta f(i) < \infty. \tag{14.7}$$

Writing $||U||$ for the L^β-norm of a random variable U, namely,

$$||U|| = (\mathbb{E}|U|^\beta)^{1/\beta},$$

(14.4) leads to

$$||T|| \leq ||\tilde{S}_{\nu_0} - S_{n_0}|| + \sum_{1\leq j<\infty} ||U_j \mathbf{1}_{[\tau> j-1]}||. \tag{14.8}$$

Conditionally given S_{n_0}, $||\tilde{S}_{\nu_0} - S_{n_0}||^\beta \leq S_{n_0}\mathbb{E}Y_1^\beta$, by Lemma 10(ii). Hence,

$$||\tilde{S}_{\nu_0} - S_{n_0}||^\beta \leq \mathbb{E}S_{n_0}\mathbb{E}Y_1^\beta < \infty.$$

Thus we need to prove the finiteness of the sum on the right of (14.8). For this define the filtration of σ-fields

$$\begin{aligned}\mathcal{G}_{2n+1} &= \sigma\{Y_j, \tilde{Y}_k : j \leq \nu_{2n+1}, k \leq \nu_{2n} + n_0\},\\ \mathcal{G}_{2n} &= \sigma\{Y_j, \tilde{Y}_k : j \leq \nu_{2n-1} + n_0, k \leq \nu_{2n}\},\ n \geq 0.\end{aligned} \tag{14.9}$$

Assume $j = 2n + 2$, $n \geq 0$. Note that

$$\begin{aligned}U_{2n+2} &= \tilde{S}_{\nu_{2n+2}} - S_{\nu_{2n+1}}\\ &= S_{\nu_{2n+1}+n_0} - S_{\nu_{2n+1}} + \tilde{S}_{\nu_{2n+2}} - S_{\nu_{2n+1}+n_0}\\ &= S_{\nu_{2n+1}+n_0} - S_{\nu_{2n+1}} + B_{2n+2}.\end{aligned}$$

Note that $S_{\nu_{2n+2}+n_0} - S_{\nu_{2n+1}}$ is independent of $\mathcal{G}_{2n+1}$, so that

$$\mathbb{E}((S_{\nu_{2n+1+n_0}} - S_{\nu_{2n+1}})^\beta|\mathcal{G}_{2n+1}) = \mathbb{E}(\sum_{k=1}^{n_0} Y_k)^\beta.$$

Hence, writing C for a generic constant, one has

$$\begin{aligned}\mathbb{E}(U_j^\beta|\mathcal{G}_{j-1}) &= \mathbb{E}(U_{2n+2}^\beta|\mathcal{G}_{2n+1})\\ &= \mathbb{E}((\tilde{S}_{\nu_{2n+2}} - S_{\nu_{2n+1}})^\beta|\mathcal{G}_{2n+1})\\ &\leq \mathbb{E}((\tilde{S}_{\nu_{2n+1}+n_0} - \tilde{S}_{\nu_{2n}} + B_{2n+2})^\beta|\mathcal{G}_{2n+1})\\ &\leq C + C\mathbb{E}((B_{2n+2})^\beta|\mathcal{G}_{2n+1}),\end{aligned} \tag{14.10}$$

using independence of $\tilde{S}_{\nu_{2n+1}+n_0} - \tilde{S}_{\nu_{2n}}$ and $\mathcal{G}_{2n+1}$. For the other term on the right side of the inequality, note that

$$\begin{aligned}B_{2n+2} &= \tilde{S}_{\nu_{2n+2}} - S_{\nu_{2n+1}+n_0}\\ &= \tilde{S}_{\nu_{2n+2}} - \tilde{S}_{\nu_{2n+1}+n_0} - (S_{\nu_{2n+1}+n_0} - \tilde{S}_{\nu_{2n}+n_0})\end{aligned} \tag{14.11}$$

This is the overshoot of a zero-delayed process $\tilde{Y}_k$, starting at $k = \nu_{2n+1} + n_0$, at the level $S_{\nu_{2n+1}+n_0} - \tilde{S}_{\nu_{2n}+n_0}$, given $\mathcal{G}_{2n+1}$. Therefore, by Lemma 10(ii),

$$\mathbb{E}\left(B_{2n+2}^\beta|\mathcal{G}_{2n+1}\right) \leq \left(\mathbb{E}Y_1^\beta\right)\mathbb{E}((S_{\nu_{2n+1}+n_0} - \tilde{S}_{\nu_{2n}+n_0})|\mathcal{G}_{2n+1}),$$

so that,

$$\begin{aligned}\mathbb{E}(U^{\beta}_{2v+2}|\mathcal{G}_{2n+1}) &\le C + C\mathbb{E}Y_1^{\beta}\mathbb{E}(S_{\nu_{2n+1}+n_0} - \tilde{S}_{\nu_{2n}+n_0}|\mathcal{G}_{2n+1})\\ &= C + C\mathbb{E}Y_1^{\beta}\mathbb{E}(B_{2n+1}|\mathcal{G}_{2n+1}).\end{aligned}$$

For,

$$S_{\nu_{2n+1}+n_0} - \tilde{S}_{\nu_{2n}+n_0} = B_{2n+1} + S_{\nu_{2n+1}+n_0} - S_{\nu_{2n+1}},$$

and

$$\mathbb{E}(S_{\nu_{2n+1}+n_0} - S_{\nu_{2n+1}}|\mathcal{G}_{2n+1}) = \mathbb{E}(S_{\nu_{2n+1}+n_0} - S_{\nu_{2n+1}}) = 0,$$

by independence. An entirely analogous argument holds for $j = 2n + 1$. Thus, we have

$$\mathbb{E}(U_j^{\beta}|\mathcal{G}_{j-1}) \le C + CB_{j-1} \text{ for all } j, \tag{14.12}$$

so that

$$\begin{aligned}\mathbb{E}(U_j^{\beta}\mathbf{1}_{[\tau>j-1]}) &= \mathbb{E}(U_j^{\beta}\mathbf{1}_{[\tau\ge j]})\\ &\le C\mathbb{E}\mathbf{1}_{[\tau\ge j]}) + C\mathbb{E}(B_{j-1}\mathbf{1}_{[\tau\ge j]}),\ (j \ge 1).\end{aligned} \tag{14.13}$$

Now

$$P(\tau \ge j|\mathcal{G}_{j-1}) = P(\tau > j-1|\mathcal{G}_{j-1}) \le (1-\gamma)^{j-1},$$

by induction on j as follows. For it holds for $j = 1, 2$ since $P(B_1 = 0|\mathcal{G}_0) > \gamma$, by the choice of n_0. Therefore,

$$P(\tau > 1|\mathcal{G}_0) = P(B_1 \ne 0|\mathcal{G}_0) < 1 - \gamma.$$

Suppose the inequality holds for j

$$P(\tau \ge j|\mathcal{G}_{j-1}) \equiv P(\tau > j-1|\mathcal{G}_{j-1}) \le (1-\gamma)^{j-1}.$$

Since $P(B_{j+1} \ne 0|\mathcal{G}_j) < 1 - \gamma$, one has

$$\begin{aligned}P(\tau \ge j+1|\mathcal{G}_j) &\equiv P(\tau > j|\mathcal{G}_j)\\ &= \mathbb{E}(\mathbf{1}_{[\tau>j-1]}\mathbf{1}_{[B_{j+1}\ne 0]}|\mathcal{G}_j)\\ &= \mathbb{E}(\mathbf{1}_{[\tau>j-1]}\mathbb{E}(\mathbf{1}_{[B_{j+1}\ne 0]}|\mathcal{G}_j)\\ &< \mathbb{E}(\mathbf{1}_{[\tau>j-1]}(1-\gamma)) < (1-\gamma)^j,\end{aligned}$$

completing the induction argument. Thus to prove that $\mathbb{E}T^{\beta} < \infty$, we need to show that

$$\sum_{j \geq 1} \mathbb{E}(B_{i-1}\mathbf{1}_{[\tau \geq j]}) < \infty. \tag{14.14}$$

We have earlier used the fact $[\tau \geq j] = [\tau > j-1] \in \mathcal{G}_{j-1}$. Therefore, arguing as in (14.11), but using Lemma 10(i), with $\rho = 1/2$,

$$\begin{aligned}
&\mathbb{E}(B_{2n+2}|\mathcal{G}_{2n+1}) \\
&= \mathbb{E}((\tilde{S}_{\nu_{2n+2}} - \tilde{S}_{\nu_{2n}+n_0}) - (S_{\nu_{2n+1}+n_0} - \tilde{S}_{\nu_{2n}+n_0}))|\mathcal{G}_{2n+1}) \\
&\leq C + 1/2B_{2n+1}, \;\; n \geq 0.
\end{aligned}$$

One may similarly estimate $\mathbb{E}(B_{2n+1}|\mathcal{G}_{2n})$ and arrive at

$$\mathbb{E}(B_{j-2}|\mathcal{G}_{j-1}) \leq C + 1/2B_{j-2}, \;\; j \geq 3. \tag{14.15}$$

Using this, we have, for $j \geq 3$,

$$\begin{aligned}
\mathbb{E}(B_{j-1}\mathbf{1}_{[\tau \geq j]}) &= \mathbb{E}(\mathbb{E}(B_{j-1}|\mathcal{G}_{j-2})\mathbf{1}_{[\tau \geq j-1]}) \\
&\leq \mathbb{E}((C + 1/2B_{j-2})\mathbf{1}_{[\tau \geq j-1]}) \\
&\leq C(1-\gamma)^{j-1} + 1/2\mathbb{E}(B_{j-2}\mathbf{1}_{[\tau \geq j-1]}) \\
&\leq C(1-\gamma)^{j-1} + 1/2(C(1-\gamma)^{j-1} + (1/2)\mathbb{E}(B_{j-3}\mathbf{1}_{[\tau \geq j-2]}) \\
&\;\;\vdots \\
&\leq C \sum_{0 \leq k \leq j-2} (1-\gamma)^{j-2-k}(1/2)^k + (1/2)^{j-1}\mathbb{E}(B_0\mathbf{1}_{[\tau \geq j]} \\
&= C \sum_{0 \leq k \leq j-2} (1-\gamma)^{j-2-k}(1/2)^k + (1/2)^{j-1}\mathbb{E}\tilde{Y}_0 \\
&\leq C \sum_{0 \leq k \leq j-1} (1-\gamma)^{j-1-k}(1/2)^k \\
&\leq Cj(\max\{1-\gamma, 1/2\})^{j-1}.
\end{aligned}$$

■

Before proceeding further, let us state also a criterion for the finiteness of $\mathbb{E}\exp\{\rho T\}$ for some $\rho > 0$.

Theorem 14.2 (Lindvall) *Assume that for some* $\gamma > 0$, $\mathbb{E}\exp\{\gamma Y_0\} < \infty$, $\mathbb{E}\exp\{\gamma \tilde{Y}_0\} < \infty$, $\mathbb{E}\exp\{\gamma Y_1\} < \infty$. *Then there exists* $\rho > 0$ *such that* $\mathbb{E}(\exp\{\rho T\}) < \infty$.

Proof We have, for $\rho > 0$, by the monotone convergence theorem,

$$\begin{aligned}\mathbb{E}\exp\{\rho T\} &\le \mathbb{E}\exp\{\rho(\tilde{S}_{\nu_0} - S_{n_0} + \sum_{j>0} U_j \mathbf{1}_{[\tau>j]})\} \\ &= \lim_{n\to\infty} \mathbb{E}\exp\{\rho((\tilde{S}_{\nu_0} - S_{n_0}) + \sum_{0<j\le n} U_j \mathbf{1}_{[\tau>j]})\}. \qquad (14.16)\end{aligned}$$

Telescoping, one can express the last expectation as

$$\begin{aligned}&\mathbb{E}\exp\{\rho((\tilde{S}_{\nu_0} - S_{n_0}) + \sum_{0<j\le n} U_j\ \mathbf{1}_{[\tau>j]})\} \\ &= 1 + \sum_{0<j\le n-1} \mathbb{E}(\exp\{\rho(\tilde{S}_{\nu_0} - S_{n_0})\} \\ &\quad \times[\exp\{\sum_{0<k\le j+1} \rho U_k \mathbf{1}_{[\tau>k]}\} - \exp\{\sum_{0<k\le j} \rho U_k \mathbf{1}_{[\tau>k]}\}]. \qquad (14.17)\end{aligned}$$

The jth term in the sum above is no greater than

$$\begin{aligned}&\mathbb{E}(\exp\{\rho(\tilde{S}_{\nu_0} - S_{n_0})\}\exp\{\sum_{0<k\le j+1} \rho U_k\}\mathbf{1}_{[\tau>j]}) \\ &\le (\mathbb{E}\exp\{2\rho(\tilde{S}_{\nu_0} - S_{n_0})\})^{1/2}(\mathbb{E}\mathbf{1}_{[\tau>j]}\exp\{2\sum_{0<k\le j+1} \rho U_k\})^{1/2}, \qquad (14.18)\end{aligned}$$

by Cauchy-Schwarz inequality. The first factor in the last line may be bounded by a constant, using the same kind of conditional argument (given S_{n_0}), and Lemma 10(iii) as used in the proof of Theorem 14.1 if $\rho > 0$ is chosen suitably. Also, using the same argument as used following (14.10)–(14.15), as well as Lemma 10(iii), one obtains, for j odd (an entirely analogous procedure holds for j even),

$$\begin{aligned}&\mathbb{E}[\exp\{2(\sum_{0<k\le j+1} \rho U_k)\}|\mathcal{G}_j] \\ &= \exp\{2\sum_{0<k\le j} \rho U_k\}\mathbb{E}[\exp\{2\rho U_{j+1}\}|\mathcal{G}_j] \\ &= \exp\{2\sum_{0<k\le j} \rho U_k\}\mathbb{E}(\exp\{2\rho(\tilde{S}_{\nu_{j+1}} - S_{\nu_j})\}|\mathcal{G}_j)\end{aligned}$$

$$
\begin{aligned}
&\leq \exp\{2 \sum_{0<k\leq j} \rho U_k\} C\mathbb{E}(\exp\{2\rho B_{j+1}\}|\mathcal{G}_j) \\
&\leq C \exp\{2 \sum_{0<k\leq j} \rho U_k\}(\mathbb{E} \exp\{2\rho Y_1\}(C + 1/2 B_{j-1}).
\end{aligned}
$$

Continuing in this way, one arrives at, just as in the proof of Theorem 14.1,

$$
\begin{aligned}
&\mathbb{E}(\exp\{2 \sum_{0<k\leq j+1} \rho U_k\})\mathbf{1}_{[\tau>j]} \\
&\leq \\
&\vdots \\
&\leq C\mathbb{E}(\exp\{2\rho Y_1\})^j (\max\{1/2, 1-\gamma\})^j. \qquad (14.19)
\end{aligned}
$$

Now choose ρ small enough such that

$$
\mathbb{E}(\exp\{2\rho Y_1\}) < (\max\{1/2, 1-\gamma\})^{-1}.
$$

Then the sum of (14.18) over $j = 0, 1, \ldots,$ converges. ■

To apply the above results to Markov processes, without much loss of generality, we take the state space S to be a Polish space and $\mathcal{S}$ its Borel σ-field. We will construct, by means of iterations of i.i.d. random maps[5] $\alpha_n, n \geq 1$, a Markov process $\{X_n : n = 0, 1, \ldots\}$ having a given transition probability $p(x, dy)$, $x \in S$. Denote by $P_\mu(F)$ probabilities of events F determined by $\{X_n\}$ with X_0 having a distribution μ, and by $P_x(F)$, if $X_0 = x$ (i.e., if $\mu = \delta_x$). Let $\mathbb{E}_\mu$, $\mathbb{E}_x$ denote corresponding expectations of random variables. Assume that there exists a set $A_0 \in \mathcal{S}$ and a probability measure ϕ, $\phi(A_0) = 1$, such that

$$
P_x(X_n \in A_0 \text{ for some } n > 0) = 1 \text{ for all } x \in S, \qquad (14.20)
$$

and there exists a $0 < \lambda < 1$ such that

$$
p(x, B) \geq \lambda\phi(B) \text{ for all } B \in \mathcal{S}, \text{ for all } x \in A_0. \qquad (14.21)
$$

Convergence to equilibrium in total variation distance for such processes, under the additional condition (14.20), is well-known and can be found in Bhattacharya and Waymire (2022), for example. To construct $\{X_n\}$ with transition probability $p(x, dy)$, we will use the decomposition of the transition probability as

[5] See, e.g., Bhattacharya and Waymire (2022), Chapter 18, Theorem 18.1, for the existence of such random maps.

$$p(x, B) = \lambda\phi(B) + (1-\lambda)q(x, B)$$
$$q(x, B) := (1-\lambda)^{-1}[p(x, B) - \lambda\phi(B)],\ x \in S, B \in \mathcal{S}. \tag{14.22}$$

Then $q(x, dy)$ is a transition probability on $(S, \mathcal{S})$. Informally, (14.22) says that we construct a Markov process $\{X_n\}$ such that given X_n, with probability λ, X_{n+1} has the distribution ϕ, and with probability $1-\lambda$, X_{n+1} has the distribution $q(X_n, dy)$. To construct $\{X_n\}$, consider an augmented probability space $(\Omega, \mathcal{F}, P)$ on which is defined an i.i.d. Bernoulli sequence $\theta_n, n \geq 1$, with

$$P(\theta_n = 1) = \lambda, \quad P(\theta_n = 0) = 1 - \lambda.$$

Also, $\{Z_n : n \geq 1\}$ is an i.i.d. sequence independent of $\{\theta_n\}$, having the common distribution ϕ on $(S, \mathcal{S})$. Finally, let X_0 be an S-vaued random variable independent of both sequences $\{\theta_n\}, \{Z_n\}$. Now define an i.i.d. sequence of random maps $\alpha_n, n \geq 1$, on S into S, with common distribution given by

$$P(\alpha_n(x) \in B) = q(x, B)\mathbf{1}_{A_0}(x) + p(x, B)\mathbf{1}_{S\setminus A_0}(x),\ B \in \mathcal{S}, x \in S. \tag{14.23}$$

By (14.22), $P(\alpha_n(x) \in B) = p(x, B)$. Now define recursively,

$$X_{n+1} = [\theta_n Z_{n+1} + (1-\theta_n)\alpha_{n+1}(X_n)]\mathbf{1}_{A_0}(X_n) + \alpha_{n+1}(X_n)\mathbf{1}_{S\setminus A_0}(X_n),\ n \geq 0. \tag{14.24}$$

It is assumed that

$$\sup_{x\in A_0} \mathbb{E}_x\tau_{A_0} < \infty, \quad \tau_{A_0} = \inf\{n \geq 1 : X_n \in A_0\},$$
$$P_x(X_n \in A_0 \text{ for some } n \geq 1) = 1 \text{ for all } x \in S. \tag{14.25}$$

Also define the following stopping times,

$$\eta_0 = 0, \quad \eta_{j+1} = \inf\{n > \eta_j : X_n \in A_0, \theta_n = 1\},\ j \geq 0;$$
$$\tau_0 = 0, \quad \tau_{j+1} = \inf\{n > \tau_j, X_n \in A_0\}, j \geq 0,\ (\tau_{A_0} := \tau_1).$$

Due to the hypothesis (14.25), the sequence of blocks

$$U_j = (X_{\eta_j+1}, X_{\eta_j+2}, \ldots; \eta_{j+1} - \eta_j),\ j \geq 1,$$

are i.i.d., and the Markov process $\{X_n : n \geq 1\}$, has a unique invariant probability π, which may be expressed as[6]

$$\pi(B) = \frac{\mathbb{E}_\phi \sum_{\eta_1 < m \leq \eta_2} \mathbf{1}_{[X_m \in B]}}{\mathbb{E}(\eta_2 - \eta_1)}. \tag{14.26}$$

Now let $\{\tilde{X}_n : n \geq 0\}$ be another Markov process with transition probability $p(x, dy)$, constructed independently of $\{X_n\}$. That is, $\{(\tilde{\theta}_n, \tilde{Z}_n, \tilde{X}_n) : n \geq 1\}$ is an independent copy of $\{(\theta_n, Z_n, X_n) : n \geq 1\}$. Also, let $\tilde{X}_0$ be independent of all these, with a distribution π. Then the integer-valued sequences $\{\eta_0, \eta_n - \eta_{n-1}\}_{n\geq 1}$ and $\{\tilde{\eta}_0, \tilde{\eta}_n - \tilde{\eta}_{n-1}\}_{n\geq 1}$, are independent and satisfy the hypotheses of Theorem 14.1 (Exercise 1). For defining stopping times, we will consider the filtration

$$\mathcal{F}_j = \sigma\{(\theta_n, Z_n, X_n), (\tilde{\theta}_n, \tilde{Z}_n, \tilde{X}_n), 1 \leq n \leq j\}, \quad \mathcal{F}_0 = \sigma\{X_0, \tilde{X}_0\}.$$

Denote again the coupling time of these renewal sequences by T. Then $X_T \in A_0$, $\theta_T = 1$, $\tilde{X}_T \in A_0$ and $\tilde{\theta}_T = 1$. Switching the X process to the $\tilde{X}$ process from time T onward, the coupling is completed. Applying Theorem 14.1, we have now arrived at the following result.

Theorem 14.3 *With the notation as above, consider a Markov process $\{X_n\}$ on $(S, \mathcal{S})$ with transition probability satisfying (14.21) and (14.25) and having initial distribution μ. (a) If, in addition, there exists $r \geq 1$ such that $\mathbb{E}_\mu \eta_1^r < \infty$, $\mathbb{E}_\pi \eta_1^r < \infty$, and $\mathbb{E}(\eta_2 - \eta_1)^r < \infty$, then $\mathbb{E}T^r < \infty$, and one has*

$$||P_\mu \circ X_n^+ - P_\pi \circ X_n^+|| = o(n^{-r}), \text{ as } n \to \infty,$$

where $P_\mu \circ X_n^+$, $P_\pi \circ X_n^+$ are the distributions of the after-n process X_n^+, with initial distributions μ and π, respectively. In particular,

$$\sup_{B \in \mathcal{S}} |P_\mu(X_n \in B) - P_\pi(X_n \in B)| = o(n^{-r}), \text{ as } n \to \infty. \tag{14.27}$$

(b) If there exists $\gamma > 0$ such that $\mathbb{E}_\mu \exp\{\gamma \eta_1\} < \infty$, $\mathbb{E} \exp\{\gamma(\eta_2 - \eta_1)\} < \infty$, $\mathbb{E}_\pi \exp\{\gamma \eta_1\} < \infty$, then there exists $\rho > 0$ such that

$$\begin{aligned} &||P_\mu \circ X_n^+ - P_\pi \circ X_n^+|| \leq C \exp\{-\rho n\}, \\ &\textit{and} \\ &\sup_{B \in \mathcal{S}} |P_\mu(X_n \in B) - P_\mu(X_n \in B)| \leq C \exp\{-\rho n\}, \ (n \geq 1). \end{aligned} \tag{14.28}$$

[6] See Bhattacharya and Waymire (2022), Proposition 20.3, Theorem 20.4.

Remark 14.2 One often uses the term strong aperiodicity referring to (14.21) holding for the one-step transition probability $p(x, dy)$ rather than a more general N-step transition probability $p^{(N)}(x, dy)$, with $N > 1$. As in the case of countable state Markov chains, one can recast the results in the general case for the (skeleton) Markov processes $\{X_{r+kN} : k = 0, 1, \ldots\}$, and average the results over N sequences for $r = 0, 1, 2, \ldots, N-1$. If the evolution is non-periodic, e.g., if (14.21) holds for both $p^{(nN)}$and $p^{(nN+1)}$, in place of p, under additional conditions such as finiteness of moments, etc., then the results of this chapter, such as convergence in total variation, and rates of convergence still hold for $p^{(nN)}$, as $n \to \infty$. It is a well known result independently obtained using coupling by Athreya and Ney (1978) and Nummelin (1978) that, under (14.21), (14.25), there exists a unique invariant probability π and that the n-step probability $p^{(n)}(x, dy)$ converges in total variation distance to π, for every $x \in S$, as $n \to \infty$.[7]

The following proposition provides a more direct way to check the moment conditions, and exponential convergence criterion, in Theorem 14.3.

Proposition 14.4 *Assume (14.21) and (14.25).*

a. *If, in addition, for some $r \geq 1$ one has*

$$c(r) := \sup_{x \in A_0} \mathbb{E}_x \tau_{A_0}^r < \infty,$$

and

$$\mathbb{E}_\mu \tau_{A_0}^r < \infty,$$

then

$$\mathbb{E}_\mu \eta_1^r < \infty, \ \mathbb{E}_\pi \eta_1^{r-1} < \infty, \ \mathbb{E}(\eta_2 - \eta_1)^r < \infty,$$

and (14.27) holds with $r - 1$, in place of r. If one also assumes $\mathbb{E}_\pi \tau_{A_0}^r < \infty$, then (14.27) holds.

b. *If*

$$C(s) := \sup_{x \in A_0} \mathbb{E}_x \exp\{s\tau_{A_0}\} < \infty, \ \mathbb{E}_\mu \exp\{s\tau_{A_0}\} < \infty,$$

for some $s > 0$, then there exists $0 < \rho \leq s$ such that

$$\mathbb{E} \exp\{\rho(\eta_2 - \eta_1)\} < \infty, \mathbb{E}_\pi \exp\{\rho\eta_1\} < \infty,$$

and (14.28) holds.

[7] See Bhattacharya and Waymire (2022), Proposition 20.3, Theorem 20.4.

Proof (a) Let us first show that $\sup_{x\in A_0}\mathbb{E}_x\eta_1^r<\infty$. Note that θ_{τ_j} is independent of τ_j, $j\geq 1$, (see (14.24), (14.25)). Therefore,

$$\begin{aligned}\mathbb{E}_x\eta_1^r &= \sum_{1\leq j<\infty}\mathbb{E}_x\tau_j^r\mathbf{1}_{[\theta_{\tau_i}=0 \text{ for } 1\leq i\leq j-1,\theta_{\tau_j}=1]},\\ &=\lambda\sum_{1\leq j<\infty}\mathbb{E}_x\tau_j^r\mathbf{1}_{[\theta_{\tau_i}=0 \text{ for } 1\leq i\leq j-1]}.\end{aligned}\tag{14.29}$$

The summand for $j=1$ is $\lambda\mathbb{E}_x\tau_{A_0}^r$. On A_0 it is bounded by $c(r)$. For $j\geq 2$, one has

$$\begin{aligned}&\mathbb{E}_x\tau_j^r\mathbf{1}_{[\theta_{\tau_i}=0 \text{ for } 1\leq i\leq j-1]}\\ &=j^r\mathbb{E}_x[j^{-1}(\sum_{1\leq k\leq j}(\tau_k-\tau_{k-1}))^r\mathbf{1}_{[\theta_{\tau_i}=0] \text{ for } 1\leq i\leq j-1]}]\\ &\leq j^r\mathbb{E}_x[j^{-1}(\sum_{1\leq k\leq j}(\tau_k-\tau_{k-1})^r)\mathbf{1}_{[\theta_{\tau_i}=0 \text{ for } 1\leq i\leq j-1]}]\\ &=j^{r-1}[(1-\lambda)^{j-1}\mathbb{E}_x(\tau_{A_0}^r)\\ &\qquad+\sum_{2\leq k\leq j}\mathbb{E}_x(\tau_k-\tau_{k-1})^r\mathbf{1}_{[\theta_{\tau_i}=0 \text{ for } 1\leq i\leq k-1]}P(\theta_{\tau_i}=0 \text{ for } k<i\leq j-1)]\\ &=j^{r-1}[(1-\lambda)^{j-1}\mathbb{E}_x(\tau_{A_0}^r)+\sum_{2\leq k\leq j}\mathbb{E}_x(\tau_k-\tau_{k-1})^r\mathbf{1}_{[\theta_{\tau_i}=0 \text{ for } 1\leq i\leq k-1]}(1-\lambda)^{j-k}]\\ &=j^{r-1}(1-\lambda)^{j-1}\mathbb{E}_x(\tau_{A_0}^r)\\ &\quad+j^{r-1}\sum_{2\leq k\leq j}(1-\lambda)^{j-k}\mathbb{E}_x(\tau_k-\tau_{k-1})^r\mathbf{1}_{[\theta_{\tau_i}=0 \text{ for } 1\leq i\leq k-1]}\\ &=j^{r-1}(1-\lambda)^{j-1}\mathbb{E}_x(\tau_{A_0}^r)\\ &\quad+j^{r-1}\sum_{2\leq k\leq j}(1-\lambda)^{j-k}\mathbb{E}_x(\mathbf{1}_{[\theta_{\tau_i}=0 \text{ for } 1\leq i\leq k-1]}\mathbb{E}((\tau_k-\tau_{k-1})^r|\mathcal{F}_{\tau_{k-1}}))\\ &\leq j^{r-1}(1-\lambda)^{j-1}c(r)+j^{r-1}\sum_{2\leq k\leq j}(1-\lambda)^{j-k}c(r)\mathbb{E}_x\mathbf{1}_{[\theta_{\tau_i}=0 \text{ for } 1\leq i\leq k-1]}\\ &=j^{r-1}(1-\lambda)^{j-1}c(r)+j^{r-1}\sum_{2\leq k\leq j}(1-\lambda)^{j-k}c(r)(1-\lambda)^{k-1}\\ &=c(r)(1-\lambda)^{j-1}j^r.\end{aligned}\tag{14.30}$$

Therefore,

$$\sup_{x\in A_0}\mathbb{E}_x\eta_1^r<\infty.$$

This implies

$$\mathbb{E}_\phi \eta_1^r = \mathbb{E}(\eta_2 - \eta_1)^r < \infty.$$

The proof for $\mathbb{E}_\mu \eta_1^r < \infty$ is obtained by integrating the term $\mathbb{E}_x(\tau_{A_0}^r)$ for $j = 1$ in (14.29) with respect to $\mu(dx)$, since the estimates for $j \geq 2$ in (14.30) do not depend on x. Similarly, $\mathbb{E}_\pi \eta_1^r < \infty$ holds if $\mathbb{E}_\pi \tau_{A_0}^r < \infty$. Now only assume $c(r) < \infty$, but not the finiteness of $\mathbb{E}_\pi \tau_{A_0}^r$. Then to prove $\mathbb{E}_\pi \eta_1^{r-1} < \infty$, it is enough to show $\mathbb{E}_\pi \tau_{A_0}^{r-1}$ is finite, in view of (14.29). Now, recalling $\tau_1 = \tau_{A_0}$,

$$\mathbb{E}_\pi \tau_1^{r-1} = \mathbb{E}_\phi \sum_{0 \leq m < \infty} \mathbb{E}_{X_m} \tau_1^{r-1} \mathbf{1}_{[m \leq \eta_1]} / \mathbb{E}(\eta_2 - \eta_1).$$

$$\begin{aligned}
&\mathbb{E}_\phi \sum_{0 \leq m < \infty} \mathbb{E}_{X_m} \tau_1^{r-1} \mathbf{1}_{[m \leq \eta_1]} \\
&= \mathbb{E}_\phi \sum_{0 \leq m < \infty} \mathbb{E}_{X_m} (\tau_1^{r-1} \circ X_m^+) \mathbf{1}_{[m \leq \eta_1]} \\
&= \mathbb{E}_\phi \sum_{0 \leq m < \infty} (\tau_1 - m)^{r-1} \mathbf{1}_{[m \leq \tau_1]} \\
&= \mathbb{E}_\phi \sum_{0 \leq m < \tau_1} (\tau_1 - m)^{r-1} \\
&\leq \mathbb{E}_\phi \tau_1^r \\
&\leq c(r) < \infty, \quad \mathbb{E}(\eta_2 - \eta_1) < \infty.
\end{aligned}$$

(b) Proceeding as in (14.29), (14.30), one has for $0 < \gamma \leq s$,

$$\mathbb{E}_x \exp\{\gamma \eta_1\} = \lambda \sum_{1 \leq j < \infty} \mathbb{E}_x (\exp\{\gamma \tau_j\} \mathbf{1}_{[\theta_{\tau_i} = 0 \text{ for } 1 \leq i \leq j-1]}).$$

For $j = 1$, the summand is $\mathbb{E}_x \exp\{\gamma \tau_1\}$. For $j \geq 2$, first taking conditional expectation, given $\mathcal{F}_{\tau_{j-1}}$, one has

$$\mathbb{E}_x \mathbb{E}(\exp\{\gamma(\tau_j - \tau_{j-1})\} | \mathcal{F}_{\tau_{j-1}}) \leq C(\gamma),$$

$$\begin{aligned}
&\mathbb{E}_x (\exp\{\gamma \tau_j\} \mathbf{1}_{[\theta_{\tau_i} = 0] \text{ for } 1 \leq i \leq j-1]}) \\
&= \mathbb{E}_x (\exp\{\gamma \sum_{1 \leq k \leq j} (\tau_k - \tau_{k-1})\} \mathbf{1}_{[\theta_{\tau_i} = 0 \text{ for } 1 \leq i \leq j-1]}
\end{aligned}$$

$$
\begin{aligned}
&\leq C(\gamma)\mathbb{E}_x(\exp\{\gamma \sum_{1\leq k\leq j-1}(\tau_k-\tau_{k-1})\}\mathbf{1}_{[\theta_{\tau_i}=0 \text{ for } 1\leq i\leq j-1]})\\
&= C(\gamma)(1-\lambda)\mathbb{E}_x(\exp\{\gamma \sum_{1\leq k\leq j-1}(\tau_k-\tau_{k-1})\}\mathbf{1}_{[\theta_{\tau_i}=0] \text{ for } 1\leq i\leq j-2]})\\
&\vdots\\
&\leq (1-\lambda)^{j-1}C(\gamma)^{j-1}\mathbb{E}_x\exp\{\gamma\tau_1\}.
\end{aligned}
$$

Note that,

$$
\mathbb{E}_x(\exp\{\gamma\eta_1\}) \leq \lambda(1+\sum_{j=2}^{\infty}(1-\lambda)^{j-1}C(\gamma)^{j-1})\mathbb{E}_x(\exp\{\gamma\tau_1\}).
$$

$C(\gamma)\downarrow 1$ as $\gamma\downarrow 0$, and we can pick γ such that $(1-\lambda)C(\gamma)<1$. This implies that

$$
\mathbb{E}_x\exp\{\gamma\eta_1\}\leq C\mathbb{E}_x\exp\{\gamma\tau_1\},\ \text{ for all } x\in S.
$$

So one needs to show that $\mathbb{E}_\pi f = \int_S f(x)\pi(dx) < \infty$, where $f(x) = \mathbb{E}_x\exp\{\gamma\tau_1\}$. As in the proof of part (a),

$$
\mathbb{E}_\pi f = \mathbb{E}_\phi \sum_{0\leq m<\infty}\mathbb{E}_{X_m}\exp\{\gamma\tau_1\}\mathbf{1}_{[m\leq\eta_1]}/\mathbb{E}(\eta_2-\eta_1),
$$

$$
\begin{aligned}
&\mathbb{E}_\phi \sum_{0\leq m<\infty}\mathbb{E}X_m\exp\{\gamma\tau_1\}\mathbf{1}_{[m\leq\eta_1]}\\
&=\mathbb{E}_\phi\sum_{0\leq m<\tau_1}\exp\{\gamma(\tau_1-m)\}\\
&=\frac{\exp\{-\gamma\}}{(1-\exp\{-\gamma\}}\mathbb{E}_\phi(\exp\{\gamma\tau_1\}-1)\\
&\leq C\mathbb{E}_\phi\exp\{\gamma\tau_1\}<\infty.
\end{aligned}
$$

■

Remark 14.3 For an irreducible aperiodic positive recurrent Markov chain (on a countable state space S), any point set $\{x_0\}$, or any finite set, may be taken as A_0. For exponential convergence to equilibrium, one well known example is the Ehrenfest model, which is a birth–death chain with two reflecting boundaries. Actually for all birth–death chains (with positive birth and death rates) on a finite set of integers with two reflecting boundaries, there is exponential convergence to equilibrium (Exercise 2).

Corollary 14.5 (Foster-Tweedie Criterion for Geometric Ergodicity[8]) *Let* $\{X_n\}$ *be a Markov process on a metric space* (S, d) *with Borel* σ*-field* $\mathcal{S}$*. Assume there exists a continuous function* $V : S \to [1, \infty)$*, bounded on a set* A_0 *satisfying (14.20), and that there are constants* $0 < a < 1$ *and* $b > 0$*, such that*

$$(T - I)V(x) \le -aV(x) + b\mathbf{1}_{A_0}(x), \quad (x \in S), \tag{14.31}$$

where T *is the one-step transition operator,*

$$Tf(x) = \int_S f(y)p(x, dy).$$

Then

(i) $\mathbb{E}_x \exp\{\gamma\tau_{A_0}\} \le c^{-1}V(x) + (b/c)\mathbf{1}_{A_0}(x)$*, where*

$$\gamma = -\ln(1 - a), \quad c = \inf_{x \in A_0} V(x).$$

(ii) If, in addition to the hypothesis in part (a), (14.21) and (14.25) hold, then there exists a unique invariant probability π *such that*

$$||P_x \circ X_n^+ - P_\pi \circ X_n^+|| \le (C_1 + C_2V(x)) \exp\{-\theta n\},$$

for all $x \in S$*, for some positive constants* C_1, C_2 *and* θ*.*

Proof First, note that $[n+1 \le \tau_{A_0}] = [n < \tau_{A_0}] \in \mathcal{F}_n$, and $[n \le \tau_{A_0}] \supset [n < \tau_{A_0}]$. Then

$$Y_n := \exp\{\gamma n\}V(X_n)\mathbf{1}_{[n \le \tau_{A_0}]}, \quad n = 1, 2, \ldots,$$

is a $\{\mathcal{F}_n\}$-supermartingale under P_x (for all $x \in S$), $\mathcal{F}_n := \sigma\{X_j : 0 \le j \le n\}$. That is,

$$\begin{aligned}
&\mathbb{E}_x(Y_{n+1} - Y_n|\mathcal{F}_n) \\
&= \mathbb{E}_x([\exp\{\gamma(n+1)\}V(X_{n+1})\mathbf{1}_{[n+1<\tau_{A_0}]} - \exp\{\gamma n\}V(X_n)\mathbf{1}_{[n \le \tau_{A_0}]}]|\mathcal{F}_n) \\
&\le \exp\{\gamma n\}[\exp\{\gamma\}\mathbb{E}_x(V(X_{n+1})\mathbf{1}_{[n<\tau_{A_0}]}|\mathcal{F}_n) - V(X_n)\mathbf{1}_{[n<\tau_{A_0}]}] \\
&= \exp\{\gamma n\}[\mathbf{1}_{[n<\tau_{A_0}]} \exp\{\gamma\}TV(X_n) - V(X_n))] \\
&\le 0, \quad (n \ge 1),
\end{aligned}$$

since

[8] See Meyn and Tweedie (1993), Theorem 15.0.1.

$$\exp\{\gamma\}TV(x) = (1-a)TV(x) \le V(x) \text{ for } x \in S \backslash A_0.$$

By the optional sampling theorem,[9] $Y_{n\wedge\tau_{A_0}}, n = 1, 2, \ldots$, is a supermartingale, so that

$$\begin{aligned}\mathbb{E}_x Y_{n\wedge\tau_{A_0}} &\le \mathbb{E}_x Y_{1\wedge\tau_{A_0}}\\ &= \mathbb{E}_x Y_1 = \mathbb{E}_x \exp\{\gamma\}V(X_1)\\ &= \exp\{\gamma\}TV(x). \end{aligned} \tag{14.32}$$

Letting $n \to \infty$, using Fatou's Lemma, one has

$$\begin{aligned}\mathbb{E}_x[\exp\{\gamma\tau_{A_0}\}V(X_{\tau_{A_0}})] &\le \exp\{\gamma\}TV(x)\\ &\le \exp\{\gamma\}(1-a)V(x) + b\mathbf{1}_{A_0}(x)\\ &= V(x) + b\mathbf{1}_{A_0}(x). \end{aligned} \tag{14.33}$$

Therefore,

$$\mathbb{E}_x \exp\{\gamma\tau_{A_0}\} \le c^{-1}V(x) + (b/c)\mathbf{1}_{A_0}(x), \ x \in S. \tag{14.34}$$

This proves part (i). Using the boundedness of V on A_0, we have

$$\sup_{x\in A_0} \mathbb{E}_x \exp\{\tau_{A_0}\} \le c^{-1} \sup_{x\in A_0} V(x) + \frac{b}{c} < \infty. \tag{14.35}$$

Part (ii) now follows from Proposition 14.4(b) and Theorem 14.3(b). ■

Remark 14.4 The proof is a little subtle, but also the simplest that we know. For examples of geometric ergodicity that derive from the above Corollary, we refer to Meyn and Tweedie (1993), Chapter 15. For applications to *nonlinear autoregressive* models, see Bhattacharya and Waymire (2022), 304-306.

We now turn to diffusions on $\mathbb{R}^d$, $d \ge 1$, generated by

$$L = (1/2) \sum_{1\le i,j\le d} a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{1\le i\le d} \mu_i(x)\frac{\partial}{\partial x_i}. \tag{14.36}$$

For simplicity, we assume that the coefficients of L are locally Lipschitz and $((a_{ij}(x)))$ is non-singular. The diffusion $\{X(t)\}$ generated by L is represented by the solution of the Itô equation

[9] See BCPT, Theorem 3.8, p. 61.

$$X_i(t) = X_i(0) + \int_{[0,t]} \mu_i(X(t))dX_i(t) + \int_{[0,t]} \sigma_i(X(t))dB(t) \ (1 \leq i \leq d). \tag{14.37}$$

Here $X(0) = (X_1(0), \ldots, X_d(0))$ is independent of the standard Brownian motion $B(t) = (B_1(t), \ldots, B_d(t))'$, $t \geq 0$; $\sigma(x)$ is a nonsingular $d \times d$ matrix such that

$$\sigma(x)\sigma'(x) = a(x),$$

and $\sigma_{i,\cdot}(x)$ is the ith row of $\sigma(x)$. In general the solution of (14.37) can only be obtained up to an explosion time ζ, but the conditions imposed for convergence to equilibrium will automatically ensure that there is no explosion, i.e., $\zeta = \infty$ a.s.. The transition probability density of $\{X(t)\}$ is denoted by $p(t; x, y)$. Let $0 < r_1 < r_2$ and, for $r > 0$,

$$B_r = \{x \in \mathbb{R}^d : |x| < r\},\ \overline{B}_r = \{x \in \mathbb{R}^d : |x| \leq r\},\ \partial B_r = \{x \in \mathbb{R}^d : |x| = r\}. \tag{14.38}$$

Define the following stopping times for $\{X(t)\}$:

$$\begin{aligned} &\tau_1 = \inf\{t \geq 0 : X(t) \in \overline{B}_r\},\ \tau_{2n+2} = \inf\{t > \tau_{2n+1} : X(t) \in \partial B_{r_1}\}, n \geq 0, \\ &\tau_{2n+1} = \inf\{t > \tau_{2n} : X(t) \in \overline{B}_r\},\ n \geq 0,\ \tau_0 = 0. \end{aligned} \tag{14.39}$$

The conditions we will impose will guarantee that the diffusion is recurrent (See Chapter 11). The following lemma will be useful for us. For positive integers α, a proof by induction may be found in Dynkin (1965), Vol. 2, Theorem 13.17, pp. 52-53.

Lemma 11 *Let G be a bounded open set with a C^1-smooth boundary, such as a ball or an annulus.*

$$\tau = \inf\{t \geq 0 : X(t) \in \partial G\},\ w_\alpha(x) = \mathbb{E}_x \tau^\alpha. \tag{14.40}$$

Then, for $\alpha \geq 2$,

$$Lw_\alpha(x) = -\alpha w_{\alpha-1}(x) \textit{ for } x \in G,\ \lim_{x \to \partial G} w_\alpha(x) = 0. \tag{14.41}$$

In the case $\alpha = 1$, $L\mathbb{E}_x\tau = -1$ for $x \in G$.

Proof Writing L_x for L, indicating differentiation with respect to x, one has (see Proposition 13.1)

$$\frac{\partial q}{\partial t} = L_x q(t : x, y),\ t > 0,\ x, y \in G,\ \lim_{x \to \partial G} q(t; x, y) = 0,\ P_x(\tau > t) = \int_G q(t : x, y)dy,$$

where $q(t; x, y)$ is the transition probability density of X in G, with absorption on ∂G and generator L. Therefore, on integration by parts,

$$w_\alpha(x) = -\int_{[0,\infty)} t^\alpha dP_x(\tau > t) = \alpha \int_{[0,\infty)} t^{\alpha-1} P_x(\tau > t)dt,$$

$$\begin{aligned} Lw\alpha(x) = \alpha \int_{[0,\infty)} t^{\alpha-1} LP_x(\tau > t)dt &= \alpha \int_{[0,\infty)} t^{\alpha-1}(\partial P_x(\tau > t)/\partial t)dt \\ &= -\alpha(\alpha-1)\int_{[0,\infty)} t^{\alpha-2} P_x(\tau > t)dt \\ &= -\alpha \mathbb{E}_x \tau^{\alpha-1} \\ &= -\alpha w^{\alpha-1}(x), \end{aligned}$$

using the fact that for a non-negative random variable U, $\mathbb{E}U^\beta = \beta \int_{[0,\infty)} u^{\beta-1} P(U > u)du$, if $\mathbb{E}U^\beta < \infty$. Note that for $\alpha \geq 2$, the integrals converge. For $\alpha = 1$, $w_1(x) = \mathbb{E}_x\tau$. It follows from Lemma 6, Chapter 11, that $\mathbb{E}_x\tau$ is finite. ■

Remark 14.5 Lemma 11 also holds for $1 < \alpha < 2$, but the proof is a little long (see Wasielak (2009), or Bhattacharya and Wasielak (2013)). It is not difficult to show directly that, for $x \in G$, $\mathbb{E}_x\tau^n < \infty$, for all $n \geq 1$, $x \in G$ (Exercise 3). However, by induction, starting with $\alpha = n = 1$, one may use (14.41) to prove this as well (See Chapter 15).

To derive rates of convergence to equilibrium, we use the following notation, as in Chapter 11: For a fixed $r_0 > 0$, define for $r \geq r_0$,

$$\begin{aligned} A(x) &= \frac{1}{|x|^2} \sum_{1\leq i,j\leq d} a_{ij}x_i x_j,\ C(x) = \sum_{1\leq i\leq d} a_{ii}(x),\ B(x) = 2\sum_{1\leq i\leq d} x_i\mu_i(x), \\ \overline{\beta}(r) &= \sup_{|x|=r} \frac{B(x)+C(x)-A(x)}{A(x)},\ \underline{a}(r) = \inf_{|x|=r} A(x), \\ \underline{\beta}(r) &= \inf_{|x|=r} \frac{B(x)+C(x)-A(x)}{A(x)}, \\ \overline{I}(r) &= \int_{[r_0,r]} (\overline{\beta}(u)/u)du,\ s(r) = \int_{[r_0,r]} \exp\{-\overline{I}(u)\}du, \\ \underline{I}(r) &= \int_{[r_0,r]} (\underline{\beta}(u)/u)du, \\ m(r) &= \int_{[r_0,r]} (\exp\{\overline{I}(u)\}/\underline{a}(u))du. \end{aligned} \tag{14.42}$$

Assumption 1: The matrix $a(x)$ is positive definite and locally Lipschitz. The drift vector $\mu(x)$ is locally Lipschitz.

Assumption 2: $\int_{[r_0,\infty)} \exp\{-\overline{I}(r)\}dr = \infty$, $\int_{[r_0,\infty)} (\exp\{\overline{I}(r)\}/\underline{a}(r))dr < \infty$.

It follows from Theorem 11.8, that the diffusion is positive recurrent, and has a unique invariant probability π if Assumptions 1,2 hold.

Theorem 14.6 *Let Assumptions 1,2 hold. Assume, in addition, that for some $\alpha \geq 1$,*

$$\int_{[r_0,\infty)} s(r)^{1-1/\alpha} m'(r)dr < \infty. \tag{14.43}$$

Then for $0 \leq \beta \leq \alpha$, $r_1 > r_0$, $x \in \partial B_{r_1}$,

$$w_\beta(x) := \mathbb{E}_x(\tau_{\partial B_{r_0}})^\beta$$

satisfies

$$w_\beta(x) \leq \Big(\prod_{1\leq i\leq[\beta]} (i+\beta-[\beta])\Big)\Big(2\int_{[r_0,\infty)} s(r)^{1-1/\alpha}\Big)^\beta s(|x|)^{\beta/\alpha}, \tag{14.44}$$

where $[\beta]$ denotes the integer part of β. In particular,

$$\sup_{x\in\partial B_{r_1}} \mathbb{E}_x(\tau_{\partial B_{r_0}})^\alpha < \infty. \tag{14.45}$$

Proof Choose x such that $|x| > r_0$. Write $\tau_{r_0} = \tau_{\partial B_{r_0}}$, $\tau_{r_0,N} = \tau_{r_0} \wedge \tau_{\partial B_N}$,

$$\begin{aligned} &w_\alpha(x) = \mathbb{E}_x\tau_{r_0}^\alpha,\ \mathbb{E}_x\tau_{r_0,N}^\alpha = \mathbb{E}_x(\tau_{r_0}\wedge\tau_{\partial B_N})^\alpha,\\ &w_{\alpha,N}(x) = \mathbb{E}_x(\tau_{r_0}\wedge\tau_{\partial B_N})^\alpha,\ C_{n,\alpha} = n!\Big(2\int_{[r_0,\infty)} s(r)^{1-1/\alpha}m'(r)dr\Big)^n. \end{aligned} \tag{14.46}$$

By the proof of Theorem 11.8, part (a), providing the upper bound below for $w_1(x) = \mathbb{E}_x\tau_{r_0}$, given by

$$\begin{aligned} w_1(x) &\leq 2\int_{[r_0,|x|)} \Big(\exp\{-\overline{I}(r)\}\Big(\int_{[u,\infty)} \exp\{\overline{I}(u)\}/\underline{a}(u)\Big)du\Big)dr\\ &= 2\int_{[r_0,|x|)} s'(u)\int_{[u,\infty)} m'(r)drdu\\ &= 2\Big(s(|x|)\int_{[|x|,\infty)} m'(r)dr + \int_{[r_0,|x|]} s(u)m'(u)du\Big)\\ &\leq 2\Big(s(|\overline{x}|)s(|x|)^{1/\alpha-1}\Big)\int_{[|x|,\infty)} s(u)^{1-1/\alpha}m'(u)du \end{aligned}$$

$$+s(|x|)^{1/\alpha} \int_{[r_0,|x|]} s(u)^{1-1/\alpha} m'(u) du)$$

$$= 2s(|x|)^{1/\alpha} \int_{[r_0,\infty)} s(u)^{1-1/\alpha} m'(u) du. \tag{14.47}$$

Let $\varepsilon \in (0, 1]$. Then,

$$w_\varepsilon(x) := \mathbb{E}_x \tau_{r_0}^\varepsilon \le (\mathbb{E}_x \tau_{r_0})^\varepsilon \le (2 \int_{[r_0,\infty)} s(u)^{1-1/\alpha} m'(u) du)^\varepsilon s(|x|)^{\varepsilon/\alpha}. \tag{14.48}$$

We will prove the following by induction,

$$w_{n+\varepsilon}(x) \le C_{n+\varepsilon,\alpha} s(|x|)^{\frac{n+\varepsilon}{\alpha}} \quad \text{for } n + \varepsilon \le \alpha. \tag{14.49}$$

Here

$$C_{n+\varepsilon,\alpha} := (\prod_{1 \le i \le n} (i + \varepsilon))(2 \int_{[r_0,\infty)} s(u)^{1-1/\alpha} m'(u) du)^{n+\varepsilon}. \tag{14.50}$$

Note that (14.49) has already been proved for $n = 0$, with the product in (14.50) over an empty set taken to be 1. Assume (14.49) for $n + \varepsilon$, and let $n + 1 + \varepsilon \le \alpha$. Lemma 11 and Itô's lemma (Corollary 8.5), applied to the stopping time $\tau_{r_0,N}$, imply

$$\begin{aligned}
w_{n+1+\varepsilon,N}(x) &= -\mathbb{E}_x \int_{[0,\tau_{r_0,N})} L w_{n+1+\varepsilon,N}(X(u)) du \\
&= (n + 1 + \varepsilon) \mathbb{E}_x \int_{[0,\tau_{r_0,N})} w_{n+\varepsilon,N}(X(u) du \\
&\le (n + 1 + \varepsilon) \mathbb{E}_x \int_{[0,\tau_{r_0,N})} w_{n+\varepsilon}(X(u) du \\
&\le (n + 1 + \varepsilon) C_{n+\varepsilon,\alpha} \mathbb{E}_x \int_{[0,\tau_{r_0,N})} s(|X(u)|)^{(n+\varepsilon)/\alpha} du,
\end{aligned} \tag{14.51}$$

where the last step follows from the induction hypothesis. Next by Lemma 12 below, and integration by parts, we get

$$\begin{aligned}
&\mathbb{E}_x \int_{[0,\tau_{r_0,N})} s(|X(u)|)^{(n+\varepsilon)/\alpha} du \\
&\le \mathbb{E}_x \int_{[r_{0,N},\infty)} \exp\{-\overline{I}(u)\} \int_{[r_0,u]} (s(u)^{(n+\varepsilon)/\alpha} dv) du
\end{aligned}$$

$$\le s(|x|)^{(n+1+\varepsilon)/\alpha}\int_{[|x|,\infty)} s(u)^{1-1/\alpha}m'(u)du$$
$$+s(|x|)^{(n+1+\varepsilon)/\alpha}\int_{[r_0,|x|]} s(u)^{(1-1/\alpha)}m'(u)du$$
$$\le (\int_{[r_0,\infty)} s(u)^{1-1/\alpha}m'(u)du)s(|x|)^{(n+1+\varepsilon)/\alpha}. \tag{14.52}$$

Substituting this in (14.51), and letting $N \uparrow \infty$, the induction is completed. Now take $\beta = n + \varepsilon$. ■

For Lemma 12, define for a nonnegative continuous function h on $[r_0, \infty)$,

$$F(r) = \int_{[r_0,r]} \exp\{-\overline{I}(u)\}\int_{[r_0,u]} (h(v)/\underline{a}(v))\exp\{\overline{I}(v)\}dv)du,\ r_0 \le r < \infty,$$

and for a positive integer N,

$$F(r;N) = \begin{cases} F(r) & \text{for } r_0 \le r \le N, \\ 0 & \text{for } r \ge 2N, \end{cases}$$

such that $F(r;N)$ is twice continuously differentiable on $[r_0,\infty)$. Also write $h(r;N) = h(r)$ for $r_0 \le r \le N$, and continuous on $[r_0,\infty)$. The following result may be found in Bhattacharya and Ramasubramanian (1982), Lemma 2.1(iv).

Lemma 12 *One has for* $x \in \partial B_{r_1}$,

$$\mathbb{E}_x\int_{[0,\tau_{r_0}]} h(|X(s)|)ds \le 2\int_{[r_0,\infty)} \exp\{-\overline{I}(u)\}\int_{[r_0,u]} (h(v)/\underline{a}(v))\exp\{\overline{I}(v)\}dv)du,$$

if the integral on the right converges.

Proof Consider the function $f(x;N) = F(|x|;N)$, and note that (See Chapter 11, Lemma 8)

$$2Lf(x;N) = A(x)F''(|x|;N) + (B(x) - A(x) + C(x))F'(|x|;N)/|x|.$$

One also has for $r_0 \le r \le N$,

$$F'(r;N) = \exp\{-\overline{I}(r)\}\int_{[r_0,r]} (h(v)/\underline{a}(v))\exp\{\overline{I}(v)\}dv) \ge 0,$$
$$F''(r;N) = -(\overline{\beta}(r)/r)\exp\{-\overline{I}(r)\}\int_{[r_0,r]} (h(v)/\underline{a}(v))\exp\{\overline{I}(v)\}dv + h(r)/\underline{a}(r)$$
$$= -(\overline{\beta}(r)/r)F'(r;N) + h(r)/\underline{a}(r),$$
$$\underline{a}(r)(F''(r;N) + (\overline{\beta}(r)/r)F'(r;N)) = h(r). \tag{14.53}$$

Therefore (Exercise 4),

$$2Lf(x; N) \geq h(|x|; N) \text{ for } |x| \leq N.$$

By Itô's lemma (Corollary 8.5), $f(X(t); N) - \int_{[0,t]} Lf(X(s); N)ds, t \geq 0$, is a martingale, so that

$$\begin{aligned}\mathbb{E}_x \int_{[0,t]} h(|X(s); N)|ds &\leq 2 \int_{[0,t]} \mathbb{E}_x Lf(X(s); N)ds \\ &= 2(\mathbb{E}_x F(|X(t); N)|) - F(|x|)).\end{aligned} \tag{14.54}$$

By the optional sampling theorem (Chapter 1), one may take $t = \tau_{\partial B_N \wedge N}$ in (14.54), and then let $N \uparrow \infty$, using the monotone convergence theorem, to get the desired result. ■

Our next task is to consider the discrete time Markov process $\{X(n) : n = 0, 1, \ldots\}$, i.e., consider the skeleton of the process $\{X(t) : t \geq 0\}$ at times $t = n = 0, 1, 2, \ldots$. Its one-step transition probability is $p(x, dy) := p(1; x, y)dy$. Let $0 < r < r_1$, and $A_0 = B_{r_2}$. Note that

$$\begin{aligned}&p(x, B) \geq \lambda\varphi(B) \text{ for all } x \in B_{r_1}, B \in \mathcal{B}(\mathbb{R}), \\ &\varphi(B) := m(B \cap \overline{B}_{r_1})/m(\overline{B}_{r_1}),\end{aligned} \tag{14.55}$$

with m being Lebesgue measure on $\mathbb{R}$, and

$$\lambda = \min\{p(1; x, y) : x, y \in \overline{B}_{r_1}\} m(\overline{B}_{r_1}).$$

Define the following stopping time for $\{X(n) : n = 0, 1, \ldots\}$: For $0 < r < r_1$,

$$\eta_1 = \inf\{n \geq 1 : X(n) \in \overline{B}_r\}. \tag{14.56}$$

The proof of the following result is quite analogous to the proof of the corresponding result, with $\alpha = 1$, in Theorem 11.7. Recall the stopping times

$$\begin{aligned}&\tau_0 = 0, \quad \tau_1 = \inf\{t \geq 0 : |X(t)| = r\} \\ &\tau_{2i+1} = \inf\{t > \tau_{2i} : |X(t)| = r\}, \quad \tau_{2i} = \inf\{t > \tau_{2i-1} : |X(t)| = r_1\}, \ i \geq 1.\end{aligned}$$

Also, let

$$\tau'_{2i+1} = \inf\{t > \tau_{2i+1} : |X(t)| = r'\}, \ 0 < r < r' < r_1 (i \geq 0).$$

Proposition 14.7 *If for some* $0 < r_1 < r_2$, $\alpha \geq 1$, *one has*

$$C := \sup_{|x|=r_1} \mathbb{E}_x \tau_1^\alpha < \infty, \tag{14.57}$$

recalling the notation in (14.39), then

$$\sup_{x \in \overline{B}_{r'}} \mathbb{E}_x \eta_1^\alpha < \infty.$$

Proof Consider the events

$$E_i = [\tau'_{2i+1} - \tau_{2i+1} \geq 1], \quad i \geq 0.$$

Let $\delta := \min_{|x|=r_1} P_x(E_0) > 0$, by the strong Feller property, and the facts that (i) P_x has full support, starting at x, for every x, and (ii) $\{x : |x| = r\}$ is compact (see Exercise 16(a)). Let $\theta = \inf\{i \geq 0 : \mathbf{1}_{E_i} = 1\}$. Then, since $\eta_1 \leq \tau_{2\theta+2}$,

$$\begin{aligned}
\eta_1^\alpha &\leq \tau_{2\theta+2}^\alpha \\
&= (\sum_{i=0}^{\theta} (\tau_{2i+2} - \tau_{2i}))^\alpha \\
&\leq (\theta + 1)^{\alpha-1} \sum_{i=0}^{\theta} (\tau_{2i+2} - \tau_{2i})^\alpha.
\end{aligned}$$

Therefore, for all $x \in \overline{B}_{r'}$,

$$\mathbb{E}_x \eta_1^\alpha \leq \sum_{n=0}^{\infty} (1-\delta)^n (n+1)^{\alpha-1} c(\alpha) < \infty,$$

where $c(\alpha) = 2^{\alpha-1}(\sup_{|x|=r_1} \mathbb{E}_x \tau_1^\alpha + \sup_{|x|=r} \mathbb{E}_x \tau_{\partial B_{r_1}}^\alpha)$. ■

In view of Propositions 14.4(a), Theorem 14.6 and Proposition 14.7, under the assumptions of Theorem 14.6, one has as $n \to \infty$,

$$||P_\mu \circ X_n^+ - P_\pi \circ X_n^+|| = o(n^{-\alpha+1}), \tag{14.58}$$

for the discrete parameter process $\{X(n) : n \geq 0\}$. Now observe that the adjoint operator

$$T_t^* \nu(dy) := \int_S p(t; x, dy) \nu(dx)$$

is a contraction on the Banach space $\mathcal{M}$ of all finite signed measures with total variation norm (Exercise 5)

$$||\nu|| := \sup_{B \in \mathcal{B}(\mathbb{R}^d)} |\nu(B)|.$$

Therefore, with

$$\nu = \upsilon - \pi, \quad \upsilon = T^*_{t-[t]}\mu, \quad (\mu \text{ a probability measure}),$$

one has

$$\begin{aligned} ||T_t^*\mu - \pi|| &= ||T_t^*(\mu - \pi)|| \le ||T^*_{[t]}(\mu - \pi)||, \\ ||T_t^{*n}\mu - \pi|| &= ||T_t^{*n}(\mu - \pi)|| \le ||T^{*n}_{[t]}(\mu - \pi)||. \end{aligned} \tag{14.59}$$

From (14.58) and (14.59) it follows that

$$\sup_{B \in \mathcal{B}(\mathbb{R}^d)} |P_\mu(X(t) \in B) - \pi(B)| = o(t^{-\alpha+1}) \text{ as } t \to \infty.$$

Applying this to the after-t process $\{X_t^+\}$, where

$$X_t^+(s) = X(t+s), \; s \ge 0,$$

we have one of the main results of this chapter.

Theorem 14.8 (Polynomial Rates of Convergence of Diffusions to Equilibrium) *Suppose Assumptions 1,2 hold. Then the diffusion $\{X(t) : t \ge 0\}$ has a unique invariant probability π.*

a. *If, also, (14.43) holds for some $\alpha \ge 1$, then*

$$||P_x \circ X_t^+ - P_\pi \circ X_t^+|| = o(t^{-\alpha+1}), \text{ as } t \to \infty, \text{ for all } x \in \mathbb{R}^d, \tag{14.60}$$

uniformly on compact subsets of $\mathbb{R}^d$.

b. *If, in addition to the hypothesis of (a), $\mathbb{E}_\pi \tau_1^\alpha < \infty$, where τ_1 is defined in (14.39), then*

$$||P_x \circ X_t^+ - P_\pi \circ X_t^+|| = o(t^{-\alpha}), \text{ as } t \to \infty, \tag{14.61}$$

uniformly (in x) on compact sets.

If also $\mathbb{E}_\mu \tau_1^\alpha < \infty$, then P_x may be replaced by P_μ in (14.61).

Proof This follows from Theorem 14.3(a), Theorem 14.6, Proposition 14.4(a), and (14.59). ■

We finally turn to a criterion for exponential rates of convergence. As in the case of deriving polynomial rates of convergence for a diffusion {X(t)} from the

corresponding rates for its discrete skeleton $\{X(n)\}$, we will first prove the following analog of Proposition 14.7.

Proposition 14.9 *With the notation as before, assume that there exists* $\gamma > 0$ *such that*

$$\sup_{|x|=r_1} \mathbb{E}_x \exp\{\gamma\tau_1\} < \infty.$$

Then there exists $0 < \gamma' < \gamma$, *such that*

$$\sup_{x\in B_{r'}} \mathbb{E}_x \exp\{\gamma'\eta_1\}.$$

Proof With the same notation as in the proof of Proposition 14.7, for $i \geq 1$, and some $\gamma > 0$ sufficiently small,

$$\begin{aligned}\mathbb{E}_x \exp\{\gamma(\tau_{2i+2} - \tau_{2i})\} &= \mathbb{E}_x(\mathbb{E}_x \exp\{\gamma(\tau_{2i+2} - \tau_{2i})\}|\mathcal{F}_{\tau_{2i}})) \\ &\leq \sup_{|x|=r_1} \mathbb{E}_x \exp\{\gamma(\tau_{2i+2} - \tau_{2i})\} = b(\gamma), \text{ say,}\end{aligned}$$

since, by the strong Markov property, the $\mathcal{F}_{\tau_{2i}}$-distribution of $X(t), t \geq 0$, is its distribution starting at $X(\tau_{2i}) \in \{x : |x| = r_1\}$. Therefore,

$$\begin{aligned}&\mathbb{E}_x(\exp\{\gamma\sum_{i=0}^{n}(\tau_{2i+2} - \tau_{2i})\}\mathbf{1}_{\cap_{j=0}^{n-1}E_j^c\cap E_n}) \\ &= \mathbb{E}_x(\exp\{\gamma\sum_{i=0}^{n-1}(\tau_{2i+2} - \tau_{2i})\}\mathbf{1}_{\cap_{j=0}^{n-1}E_j^c})\mathbb{E}_x(\exp\{\gamma(\tau_{2n+2} - \tau_{2n})\}\mathbf{1}_{E_n}|\mathcal{F}_{\tau_{2n}}) \\ &\leq b(\gamma)\mathbb{E}_x(\exp\{\gamma\sum_{i=0}^{n-1}(\tau_{2i+2} - \tau_{2i})\}\mathbf{1}_{\cap_{j=0}^{n-1}E_j^c}).\end{aligned}$$

Now, $E_{n-1}^c \in \mathcal{F}_{\tau'_{2n-1}}$, and $\mathcal{F}_{2n-2} \subset \mathcal{F}_{\tau'_{2n-1}} \subset \mathcal{F}_{2n}$, so that

$$\begin{aligned}&\mathbb{E}_x(\exp\{\gamma(\tau_{2n} - \tau_{2n-2})\}\mathbf{1}_{E_{n-1}^c}|\mathcal{F}_{\tau_{2n-2}}) \\ &= \mathbb{E}_x(\mathbb{E}_x(\exp\{\gamma(\tau_{2n} - \tau_{2n-2})\}|\mathcal{F}_{\tau'_{2n-1}})\mathbf{1}_{E_{n-1}^c}|\mathcal{F}_{\tau_{2n-2}}).\end{aligned}$$

Note that $|X(\tau'_{2n-1})| = r'$. Therefore, the last expression reduces to

$$\begin{aligned}&\mathbb{E}(\mathbb{E}_x(\exp\{\gamma(\tau_{2n} - \tau'_{2n-1})\}|\mathcal{F}_{\tau'_{2n-1}})\exp\{\gamma(\tau'_{2n-1} - \tau_{2n-2})\}\mathbf{1}_{E_{n-1}^c}|\mathcal{F}_{\tau_{2n-2}})) \\ &\leq \mathbb{E}_x(\exp\{\gamma(\tau_{2n-1} + 1 - \tau_{2n-2})\}\mathbf{1}_{E_{n-1}^c}|\mathcal{F}_{\tau_{2n-2}}) \\ &= c_1e^{\gamma}\mathbb{E}_x(\exp\{\gamma(\tau_{2n-1} - \tau_{2n-2})\}\mathbf{1}_{E_{n-1}^c}|\mathcal{F}_{\tau_{2n-2}}),\end{aligned}$$

where $c_1 = \sup_{|x|=r'} \mathbb{E}_x \exp\{\gamma \tau_{\partial B_{r_1}}\} < \infty$, if γ is sufficiently small (Exercise 16(b)). Finally,

$$
\begin{aligned}
&\mathbb{E}_x(\exp\{\gamma(\tau_{2n-1} - \tau_{2n-2})\}\mathbf{1}_{E^c_{n-1}}|\mathcal{F}_{\tau_{2n-2}}) \\
&= \mathbb{E}_x(\mathbb{E}_x(\exp\{\gamma(\tau_{2n-1} - \tau_{2n-2})\}\mathbf{1}_{E^c_{n-1}}|\mathcal{F}_{\tau_{2n-1}})|\mathcal{F}_{\tau_{2n-2}})) \\
&= \mathbb{E}_x((\exp\{\gamma(\tau_{2n-1} - \tau_{2n-2})\})\mathbb{E}_x(\mathbf{1}_{E^c_{n-1}}|\mathcal{F}_{\tau_{2n-2}})) \\
&\leq \mathbb{E}_x(\exp\{\gamma(\tau_{2n-1} - \tau_{2n-2})\}|\mathcal{F}_{\tau_{2n-2}}) \sup_{|x|=r} P_x(\tau_{\partial B_{r'}} \leq 1) \\
&\leq (1-\delta)\mathbb{E}_x \exp\{\gamma(\tau_{2n-1} - \tau_{2n-2})\} \\
&\leq c_1 e^{\gamma} \mathbb{E}_x \sup_{|x|=r_1} \mathbb{E}_x \exp\{\gamma\tau_1\}(1-\delta).
\end{aligned}
$$

Continuing in this way, one arrives at

$$
\begin{aligned}
&\mathbb{E}_x(\exp\{\gamma \sum_{i=0}^{n}(\tau_{2i+2} - \tau_{2i})\}\mathbf{1}_{\cap_{j=0}^{n-1} E^c_j \cap E_n}) \\
&\leq (c_1 e^{\gamma})^{n-1} b(\gamma)(\sup_{|x|=r_1} \exp\{\gamma\tau_1\})^{n-1}(1-\delta)^n,
\end{aligned}
$$

where $c_1 = \sup_{|x|=r'} \mathbb{E}_x \exp\{\gamma\tau_{\partial B_{r_1}}\}$ and $b(\gamma) = \sup_{|x|=r_1} \mathbb{E}_x \exp\{\gamma(\tau_2 - \tau_0)\}$. If, instead of γ, one chooses $\gamma' < \gamma$ sufficiently small, then one would get

$$
\sup_{x \in B_{r'}} \mathbb{E}_x \exp\{\gamma'\eta_1\} = \sup_{|x| \leq r'} \mathbb{E}_x \exp\{\gamma'\eta_1\}.
$$

Note that one needs γ' such that

$$
c_1 e^{\gamma'} \sup_{|x|=r_1} \mathbb{E}_x \exp\{\gamma'\tau_1\} < (1-\delta)^{-1}.
$$

Except for c_1, the other terms converge to 1 as $\gamma' \downarrow 0$. Choosing γ' sufficiently close to γ_1, one can reduce c_1 sufficiently and, at the same time make δ large and $1-\delta$ sufficiently small. ■

Theorem 14.10 (Exponential Rates of Convergence of Diffusions to Equilibrium) *Suppose Assumptions 1 and 2 hold, and that for some $r_1 > r_0 > 0$, $\delta > 0$, $\lambda > 0$,*

$$
\frac{Trace(a(x)) - (2-\delta)A(x) + 2x\mu(x)}{|x|^2} < -\lambda \text{ for } |x| \geq r_0. \tag{14.62}
$$

Then

(i) There exists λ', $0 < \lambda' < \lambda$ *such that*

$$\sup_{x \in \partial B_{r_1}} \mathbb{E}_x \exp\{\lambda' \tau_1\} < \infty, \tag{14.63}$$

where $\tau_1 := \tau_{\partial B_{r_0}}$.

(ii) There exists $\lambda'' > 0$ *such that*

$$||P_x \circ X_t^+ - P_\pi \circ X_t^+|| = o(\exp\{-\lambda'' t\}) \text{ as } t \to \infty, \tag{14.64}$$

uniformly on compact subsets of $\mathbb{R}^d$.

If $\mathbb{E}_\mu \exp\{\lambda \tau_1\} < \infty$, *then (14.64) holds for* P_μ *in place of* P_x.

Proof Let

$$g(t, x) = \exp\{\gamma t\}|x|^\delta = \exp\{\gamma t\} g(x)$$

for some positive constant γ to be determined later. Let $\eta_M = \inf\{t \geq 0 : |X(t)| = M\}$. By Itô's Lemma,

$$\begin{aligned}
&\mathbb{E}_x[\exp\{\gamma(\tau_1 \wedge \eta_M)\} g(|X(\tau_1 \wedge \eta_M)|) - |x|^\delta \\
&= \mathbb{E}_x \int_{[0, \tau_1 \wedge \eta_M]} \exp\{\gamma s\}[\gamma |X(s)|^\delta + Lg(X(s)]ds, \; r_0 \leq |x| \leq M.
\end{aligned} \tag{14.65}$$

Note that

$$2Lg(x) = \delta |x|^{\delta - 2}[(\delta - 2)A(x) + B(x) + C(x)].$$

Hence, by (14.62),

$$2Lg(x) \leq -\lambda \delta |x|^\delta \text{ for } |x| \geq r_0.$$

Choose $\gamma = (\delta/2)\lambda$, to show that $\gamma |x|^\delta + Lg(x) \leq 0$ for $|x| \geq r_0$. Hence (14.65) yields

$$\mathbb{E}_x[\exp\{\gamma \tau_1 \wedge \eta_M\} g(|X(\tau_1 \wedge \eta_M)|)] \leq |x|^\delta \text{ for } r_0 \leq |x| \leq M.$$

Letting $M \uparrow \infty$, one obtains

$$\mathbb{E}_x \exp\{\lambda' \tau_1\} \leq r_1^\delta \text{ for } |x| = r_1,$$

for all $0 < \lambda' < \lambda$. The proof of (14.63) is completed by using Proposition 14.9 and Proposition 14.4(b). ■

Remark 14.6 In view of the results of Stroock and Varadhan (1979), the smoothness Assumption 1 may be relaxed to: (i) $a(x)$ is continuous and positive definite, (ii) the drift $\mu(x)$ is Borel measurable and bounded on compacts (See Bhattacharya (1978)). In one dimension the most general assumptions are those of Feller (see Chapter 21).

Remark 14.7 In the self-adjoint case, Chen (2006) and his collaborators have obtained sharp results on the speed of convergence in terms of the spectral gap. Also see Chen and Wang (1995), Mao (2006) in this regard.

Remark 14.8 A successful coupling of a Markov process, in the sense described in this chapter, implies convergence in total variation distance of the process (being coupled) to the one under equilibrium. There are important classes of Markov processes with unique invariant probabilities and convergence in distribution to equilibrium under every initial state, but for which the convergence is not in total variation distance (see Bhattacharya and Waymire (2022), Chapters 18 and 19). See Exercise 11.

Next let us consider some illustrative examples. Some may defy intuition.

Example 1 (Polynomial Rates of Convergence to Equilibrium for One-Dimensional Diffusions) Let $\{X(t)\}$ be a positive recurrent one-dimensional diffusion. Although Theorem 14.8 applies, one may improve it, or simplify, because of point recurrence. Take two distinct points $x_0 = 0$, x_1 and define stopping times

$$\tau_j =: \inf\{t > \tau_{j-1} : X(t) = x_0 \text{ after visiting } x_1\},\ \ j \geq 1,$$

$\tau_0 = 0$. Then $\{\tau_j - \tau_{j-1} : j \geq 2\}$ are i.i.d, and one may obtain a sufficient condition for rate of convergence $o(t^{-\alpha+1})$, $\alpha \geq 1$, of $P_x \circ X_t^+$ to equilibrium $P_\pi \circ X_t^+$ assuming $\mathbb{E}(\tau_2 - \tau_1)^\alpha < \infty$. This amounts to (14.43), by deleting the 'overbar', namely,

$$\int_{[0,\infty)} s(y)^{1-1/\alpha} m(dy) < \infty,\ \ \int_{(-\infty,0)} |s(y)|^{1-1/\alpha} m(dy) < \infty, \tag{14.66}$$

where

$$\begin{aligned} I(y) &= \int_{[0,y)} (2\mu(z)/\sigma^2(z))dz\ (y \in \mathbb{R}),\ \ s(y) = \int_{[0,y)} \exp\{-I(z)\}dz, \\ m(y) &= \int_{[0,y]} 2(\exp\{I(y)\}/\sigma^2(y))dy, \end{aligned} \tag{14.67}$$

with the usual convention, $\int_{(y,0)} = -\int_{(0,y)}$. Usually it is the case that positive recurrence occurs if the drift $\mu(y) < 0$ for large positive y, and $\mu(y) > 0$ for y negative of large magnitude. Some curious examples follow. Assume, for simplicity, the coefficients to be locally Lipschitz. Let $\mu(y) = 0$ for all y, $\sigma^2(y) > 0$ and grows fast as $|y| \to \infty$, such that

$$m(\infty) = \int_{[0,\infty)} (2/\sigma^2(y))dy < \infty, \ -m(-\infty) = \int_{(-\infty,0]} (2/\sigma^2(y))dy < \infty. \tag{14.68}$$

To see this, first note that $s(y) = y$ for all y, implying $s(\infty) = \infty$, $s(-\infty) = -\infty$. Hence the diffusion is recurrent (Theorem 8.8); in particular, there is no explosion. Next, (14.68) says that $m(\infty) < \infty$, and $m(-\infty) > -\infty$. That is, the diffusion is positive recurrent (Theorem 11.9(c) and (11.33)), with invariant probability π having the density

$$\pi'(y) = m'(y)/(m(\infty) - m(-\infty)). \tag{14.69}$$

(i) Consider the special case $\mu(\cdot) = 0$, $\sigma^2(y) = 1 + |y|^a$ for some a. Then check that the diffusion is null recurrent if $a \le 1$, and positive recurrent if $a > 1$; also, for a given $\alpha \ge 1$, the rate of convergence to equilibrium is $o(t^{-\alpha+1})$ as $t \to \infty$, if $a > 2 - 1/\alpha$ (Exercise 6).
(ii) Now consider the case $\mu(y) \equiv \theta \neq 0$, $\sigma^2(y) = 1 + |y|^a$ for some a. Then check that the diffusion is transient if $a < 1$, and positive recurrent if $a > 1$. If $\alpha \ge 1$, and $a > 2 - 1/\alpha$, then the rate convergence to equilibrium is $o(t^{-\alpha+1})$ as $t \to \infty$. What happens for $a = 1$, depends on the value of θ (Exercise 7).
(iii) Finally, note that specifications of $\mu(y)$ and $\sigma^2(y)$ in (i), (ii), are only needed for $|y| > R$, for some R, however large R may be (Exercise 12).

Thus, surprisingly, although the constant drift tends to push the diffusion to ∞ if $\theta > 0$, and toward $-\infty$ if $\theta < 0$, the large diffusion coefficient, with $a > 1$, prevents transience, and even allows convergence to a steady state!

Example 2 (Polynomial and Exponential Rates of Convergence to Equilibrium for Multi-dimensional Diffusions) First, here are analogs of Examples 1(i),(ii). Let $d > 1$, I_d denotes the $d \times d$ identity matrix. We will denote the diffusion matrix by $D(y)$. We will consider radial diffusions, although one may similarly consider non-radial diffusions with radial approximations.

(i) Let $\mu(\cdot) = 0$, $D(y) = (1 + |y|^a)I_d$. Then, using the notation (14.42) with $r_0 = 1$, $\overline{\beta}(r) = d - 1$, $\tilde{a}(r) = 1 + r^a$, $\overline{I}(r) = (d-1)\ln r$, $s(r) = \ln r$ if $d = 2$; $s(r) = (1 - r^{-(d-2)})/(d-2)$, if $d > 2$. Then, for all a, the diffusion is recurrent if $d = 2$ and transient if $d > 2$. If $a > 2$, the diffusion is positive recurrent for $d = 2$. In the latter case, the rate of convergence to equilibrium is $o(t^{-\alpha+1})$ for all $\alpha \ge 1$, in view of Theorem 14.8. That is, the convergence to equilibrium is faster than any polynomial rate (Exercise 7)! If $a \le 2$, then the two-dimensional diffusion is null recurrent.
(ii) Let $\mu(y) = \theta y$ for some non-zero θ, $D(y) = (1 + |y|^a)I_d$. In this case $\beta(r) = d - 1 + (2\theta r^2/(1 + r^a))$. Suppose $\theta > 0$. Then one may check that for all $a \le 2$, the diffusion is transient, whatever be the dimension $d \ge 2$. For $a > 2$ the diffusion is recurrent for $d = 2$, and transient for $d > 2$. In

this model the diffusion is transient for $d > 2$, whatever a is. For $d = 2$, the diffusion is transient if $a < 2$, and null recurrent if $a \geq 2$ (Exercise 8).

(iii) Next, consider the case (ii), but with $\theta < 0$. Then the diffusion is positive recurrent for all values of a. The speed of convergence to equilibrium is exponential (1) for all values of a, if $d = 1$, (2) for $a < 2$, for all d, and (3) for $a = 2$, if $|\theta| > d - 2$. This follows from Theorem 14.10 (Exercise 9).

(iv) The results (i)-(iii) also hold if the hypothesis about $\mu(y)$ and $D(y)$ hold only for $|y| > R$, however large R may be (Exercise 13).

Exercises

1. Prove that, in the notation of(14.26), the integer-valued sequences $\{\eta_0, \eta n - \eta n - 1 : n \geq 1\}$ and $\{\tilde{\eta}_0, \tilde{\eta}_n - \tilde{\eta}_{n-1} : n \geq 1\}$, with $\tilde{\eta}_0$ having the invariant distribution of the residual renewal process, are independent and satisfy the hypotheses of Theorem 14.1, with $Y_n = \eta_n - \eta_{n-1}$, $\tilde{Y}_n = \tilde{\eta}_n - \tilde{\eta}_{n-1}$, $n \geq 1$.
2. Prove that a birth–death chain on a finite set, with both boundaries reflecting, has a unique invariant probability, and the convergence to equilibrium is exponential, no matter the initial state.
3. In Remark 14.5, directly prove $\mathbb{E}_x \tau^n < \infty$, for all $n \geq 1$.
4. Check (14.53) and that $2Lf(x; N) \geq h(|x|; N)$ for $|x| \leq N$.
5. Prove that the adjoint operator T^* is a contraction on the space of finite signed measures endowed with total variation norm.
6. Check the assertions in Example 1(i), and compute the invariant distribution when $\sigma^2(y) = \exp\{|y|\}$, $\mu(y) = 0$.
7. Check the assertions in Example 1(ii), with $\mu(y) = \theta$, $\sigma^2(y) = 1 + |y|^a$. Describe the asymptotics in the case $a = 1$.
8. Check the assertions in Example 2(i).
9. Check the assertions in Example 2(ii).
10. Check the assertions in Example 2(iii).
11. Find a Markov process on an interval $[a, b]$ with a transition probability p having the Feller property, such that $p^{(n)}(x, dy)$ is supported on a countable set for every n and every $x \in [a, b]$, but has a unique absolutely continuous invariant probability.
12. Show that the specifications of $\mu(y), \sigma^2(y)$ in Example 1 are needed only for $|y| > R$, however large R may be.
13. Prove the assertion in part (iv) of Example 2.
14. In Example 1(ii), assume that $\mu(y) \to \theta \neq 0$ as $|y| \to \infty$, and that $\sigma^2(y) \sim 1 + |y|^a$ as $|y| \to \infty$. (Here $\sim$ means the ratio of the two sides goes to 1). Derive the asymptotics in this case without requiring specifications of $\mu(y)$ and $\sigma^2(y)$ for all y.
15. (*Stochastic Logistic Growth Model*) (a) Consider the one-dimensional diffusion of Example 1 in Chapter 11: $dX(t) = rX(t)(k - X(t))dt + \sigma X(t)dB(t),\ t \geq 0,$

where $r > 0, k > 0, \theta > 0$. Estimate the speed of convergence to equilibrium if $rk - \theta > \frac{\sigma^2}{2}$.

16. (a) In the proof of Proposition 14.7, show that $\delta := \inf_{|x|=r_1} P_x(\tau_1' - \tau 1 \geq 1) > 0$. [*Hint*: If $|x| = r_1$, then $P_x(\tau_1' - \tau_1 \geq 1) = P_x(X(\tau_1) < \infty, \tau_1' - \tau_1 \geq 1) = \mathbb{E}_x P_{X(\tau_1)}(\tau_{\partial B_{r'}} \geq 1) \geq \inf_{|x|=r} P_x(\tau_{\partial B_{r'}} \geq 1) > 0$, since, for $|x| = r$, $P_x(\tau_{\partial B_{r'}} \geq 1) > 0$, because (i) P_x has full support, (ii) $x \to P_x(\tau_{\partial B_{r'}} \geq 1)$ is continuous on $|x| = r$, by the strong Feller property, and (iii) the last infimum is achieved on the compact set $\{x : |x| = r\}$.] (b) In the proof of Proposition 14.9, prove that $\sup_{|x|=r'} \mathbb{E}_x \exp\{\gamma \tau_{\partial B_{r'}}\} < \infty$ for a sufficiently small $\gamma > 0$. [*hint*:$P_x(\tau_{\partial B_{r'}} > 1) < 1$ for all $x \in \{x : |x| \leq r'\}$, in view of the full support of P_x. Let $\theta := \sup_{|x|\leq r'} P_x(\tau_{\partial B_{r_1}} > 1)$. Then $\theta < 1$ by the continuity of $x \to P_x(\tau_{\partial B_{r'}} > 1)$ (strong Feller property). Also, iterating and using the strong Markov property repeatedly with stopping time $\tau_{\partial B_{r_1}}$, one has $P_x(\tau_{\partial B_{r_1}} > n) \leq \theta^n$ for all n. Hence $\mathbb{E}_x \exp\{\gamma \tau_{\partial B_{r_1}}\} \leq \sum_{n=0}^{\infty} e^{\gamma n} \theta^n < \infty$ if $e^\gamma < 1/\theta$.]

Chapter 15
Probabilistic Representation of Solutions to Certain PDEs

The theory and methods of partial differential equations (pdes) have a long history in the context of diffusions. This is natural from the more analytic perspective of semigroup theory. However, the introduction of stochastic calculus in terms of stochastic differential equations provides an alternative probabilistic approach in which to analyze pdes.

Recall from Chapter 2 that from an analytic point of view, the Markov property is partly conveyed in the *semigroup property* of the family of linear operators $\{\mathbf{T}_t : t \geq 0\}$ defined on the vector space of bounded, measurable functions f by

$$\mathbf{T}_t f(\mathbf{x}) = \mathbb{E}_{\mathbf{x}} f(X_t) = \int_{\mathbb{R}^k} f(\mathbf{y}) p(t; \mathbf{x}, d\mathbf{y}), \quad t \geq 0, \quad \mathbf{x} \in \mathbb{R}^k. \tag{15.1}$$

That is, one has

$$\mathbf{T}_{t+s} f = \mathbf{T}_t \mathbf{T}_s f, \quad s, t \geq 0.$$

Let $C_b(\mathbb{R}^k)$ be the Banach space of bounded, continuous functions on $\mathbb{R}^k$ with the uniform norm. Observe that $\mathbf{T}_t : C_b(\mathbb{R}^k) \to C_b(\mathbb{R}^k)$ is implied by the Feller property established as a result of Proposition 7.8. The necessity of $\mathbf{T}_t f - f = o(1)$ uniformly as $t \downarrow 0$, may be recast in terms of Itô's lemma as follows:

$$\mathbf{T}_t f(\mathbf{x}) - f(\mathbf{x}) = \mathbb{E}\{f(\mathbf{X}_t^{\mathbf{x}} - f(\mathbf{x})\} = \mathbb{E} \int_0^t \mathbf{A} f(X_s^{\mathbf{x}}) ds. \tag{15.2}$$

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_15

Further conditions on $f \in C_b(\mathbb{R}^k)$ generally involve bounds on norms of $\boldsymbol{\mu}(\mathbf{X}_s^{\mathbf{x}})$, $\nabla(\mathbf{X}_s^{\mathbf{x}})$, and, more generally, $\mathbf{A}f$ as follows. As noted earlier, the Lipschitz condition on $\boldsymbol{\mu}$, for example, implies

$$|\boldsymbol{\mu}(\mathbf{X}_s^{\mathbf{x}})| \leq |\mathbf{x}| + K|\mathbf{X}_s^{\mathbf{x}} - \mathbf{x}|.$$

A condition of polynomial growth of ∇f, i.e., $|\nabla f(\mathbf{x})| \leq c(1 + |\mathbf{x}|^{2m})$ for some positive constant c and nonnegative integer m, controls the size of $\nabla f(\mathbf{X}_s^{\mathbf{x}})$ in (15.2) in terms of a moment bound. In this regard, it is useful to recall Doob's L^p maximal inequality for moments of right-continuous martingales $\{Z_t : \alpha \leq t \leq \beta\}$ with $\mathbb{E}|Z_t|^p < \infty$, $p > 1$ (Theorem 1.15),

$$\mathbb{E}|\max_{\alpha \leq t \leq \beta} Z_t|^p \leq (\frac{p}{p-1})^p \mathbb{E}|Z_\beta|^p. \tag{15.3}$$

A function $f : \mathbb{R}^k \to \mathbb{R}$ is said to *vanish at infinity* if $|f(x)| \to 0$ as $|x| \to \infty$.

Proposition 15.1 *Consider the semigroup $\{\mathbf{T}_t : t \geq 0\}$ defined by (15.1) on the space $C_b(\mathbb{R}^k)$, where $\{\mathbf{X}^{\mathbf{x}} : \mathbf{x} \in \mathbb{R}^k\}$ is the diffusion defined by Lipschitzian coefficients $\boldsymbol{\mu}(\cdot), \boldsymbol{\sigma}(\cdot)$ having Lipschitz constant K. Recall the linear operator A introduced in (8.13). Namely,*

$$A = \sum_{1 \leq i \leq k} \mu^{(i)}(\mathbf{x}) \frac{\partial}{\partial x^i} + \frac{1}{2} \sum_{1 \leq i,j \leq k} a_{ij}(\mathbf{x}) \frac{\partial^2}{\partial x^i \partial x^j}, \quad D = ((a_{ij})) = \sigma\sigma^t.$$

Then the semigroup and its generator have the following properties:

1. (a) $\|\mathbf{T}_t f - f\| \to 0$ *as* $t \downarrow 0$, *for every real-valued Lipschitzian f on $\mathbb{R}^k$ vanishing at infinity. (b) If $\mu(\cdot), \sigma(\cdot)$ are bounded, then $||T_t f - f|| \to 0$ for all bounded, uniformly continuous functions f.*

2. *Let f be a real-valued twice continuously differentiable function on $\mathbb{R}^k$.*

- **a.** *If $\mathbf{A}f$ and* **grad** *f are polynomially bounded (i.e., $|\mathbf{A}f(\mathbf{x})| \leq c_1(1 + |\mathbf{x}|^{m_1})$, $|\,\mathbf{grad}\, f(\mathbf{x})| \leq c_2(1 + |\mathbf{x}|^{m_2})$, then $\mathbf{T}_t f(\mathbf{x}) \to f(\mathbf{x})$ as $t \downarrow 0$, uniformly on every bounded set of $\mathbf{x}$'s.*
- **b.** *If $\mathbf{A}f$ is bounded and* **grad** *f polynomially bounded, then $\|\mathbf{T}_t f - f\| \to 0$.*

Proof Given $\varepsilon > 0$, let $K(\varepsilon)$ be such that $|f(x)| < \varepsilon/2$ if $|x| > K(\varepsilon)$. For $|x| \leq K(\varepsilon)$, $|\mathbb{E}f(X_t^x) - f(x)| < C(\varepsilon)t$ for some constant $C(\varepsilon)$ (using Exercise 7(i) with $||X_0||_{2m} = K(\varepsilon)$); hence $|T_t f(x) - f(x)| < \varepsilon$ if $t < \varepsilon/C(\varepsilon)$. For $|x| > K(\varepsilon)$, $|f(x)| < \varepsilon/2$, so that $|T_t f(x) - f(x)| < |T_t f(x)| + |f(x)| < \varepsilon$, if $t < \varepsilon/(2||f||)$. Thus, if $t < \delta := \min\{\varepsilon/C(\varepsilon), \varepsilon/(2||f||)\}$, then $||T_t f - |f|| < \varepsilon$. For (1b), see Exercise 8. For (2a), use Itô's lemma and Exercise 7(iii). For (2b), use Exercise 6(ii). ■

Definition 15.1 (Infinitesimal Generator) Let $\mathcal{D}$ denote the set of all real-valued bounded continuous functions f on $\mathbb{R}^k$ such that $\|t^{-1}(\mathbf{T}_t f - f) - g\| \to 0$ as $t \downarrow 0$, for some bounded continuous g (i.e., $g = ((d/dt)\mathbf{T}_t f)_{t=0}$), where the convergence of the difference quotient to the derivative is uniform on $\mathbb{R}^k$. Write $g = \hat{\mathbf{A}} f$ for such f. The operator $\hat{\mathbf{A}}$ on the *domain* $\mathcal{D}$ is called the *infinitesimal generator of* $\{\mathbf{T}_t : t \geq 0\}$.

Remark 15.1 In this context, $\mathbf{A}$ is often referred to as the *formal infinitesimal generator*, according to which $\mathbf{A} f(x)$ is given by the indicated pointwise operations for all twice continuously differentiable functions f.

To determine functions in $\mathcal{D}$ together with an explicit recipe to compute $\hat{\mathbf{A}}$ one requires conditions such that

$$\mathbf{T}_t f - f - tg = o(t), \quad \text{uniformly as} t \downarrow 0.$$

Once again, this may be analyzed by an application of Itô's lemma to $f(\mathbf{X}_t^{\mathbf{x}})$, minimally requiring f to be twice continuously differentiable to get started.

Proposition 15.2 *Assume that $\mu(\cdot)$ and $\sigma(\cdot)$ are Lipschitz.*

1. *Every twice continuously differentiable f, which vanishes outside some bounded set, belongs to $\mathcal{D}$ and, for such f, $\hat{\mathbf{A}} f = \mathbf{A} f$ where $\mathbf{A}$ is given by (8.35) or in Proposition 15.1.*
2. *Let f be a real-valued, bounded, twice continuously differentiable function on $\mathbb{R}^k$ such that* **grad** *f is polynomially bounded, and $\mathbf{A} f(\mathbf{x}) \to 0$ as $|\mathbf{x}| \to \infty$. Then $f \in \mathcal{D}$ and $\hat{\mathbf{A}} f = \mathbf{A} f$.*
3. *(Local Property of $\hat{\mathbf{A}}$) Suppose $f, g \in \mathcal{D}$ are such that $f(\mathbf{y}) = g(\mathbf{y})$ in a neighborhood $B(\mathbf{x} : \varepsilon) \equiv \{\mathbf{y} : |\mathbf{y} - \mathbf{x}| < \varepsilon\}$ of $\mathbf{x}$. Then $\hat{\mathbf{A}} f(\mathbf{x}) = \hat{\mathbf{A}} g(\mathbf{x})$.*

Proof For (a) suppose $f(\mathbf{x}) = 0$ for $|\mathbf{x}| \geq R$. For each $\varepsilon > 0$ and $\mathbf{x} \in \mathbb{R}^k$, use Itô's lemma to obtain

$$|\mathbf{T}_t f(\mathbf{x}) - f(\mathbf{x}) - t\mathbf{A} f(\mathbf{x})| \leq t\theta(\mathbf{A} f, \varepsilon) + 2t\|\mathbf{A} f\| P\left(\max_{0 \leq s \leq t} |\mathbf{X}_s^{\mathbf{x}} - \mathbf{x}| \geq \varepsilon\right),$$

where $\theta(h, \varepsilon) = \sup\{|h(\mathbf{y}) - h(\mathbf{z})| : \mathbf{y}, \mathbf{z} \in \mathbb{R}^k, |\mathbf{y} - \mathbf{z}| < \varepsilon\}$. As $\varepsilon \downarrow 0$, $\theta(\mathbf{A} f, \varepsilon) \to 0$, and for each $\varepsilon > 0$,

$$P(\max_{0 \leq s \leq t} |\mathbf{X}_s^{\mathbf{x}} - \mathbf{x}| \geq \varepsilon) \to 0 \qquad \text{as } t \downarrow 0$$

uniformly for all $\mathbf{x}$ in $B_{R+\varepsilon} := \{\mathbf{y} : |\mathbf{y}| \leq R + \varepsilon\}$ (see Exercise 8(iii)). For $\mathbf{x} \in B_{R+\varepsilon}^c$,

$$|\mathbf{T}_t f(\mathbf{x}) - f(\mathbf{x}) - t\mathbf{A} f(\mathbf{x})| = |\mathbb{E}\int_0^t \mathbf{A} f(\mathbf{X}_s^{\mathbf{x}}) ds| \leq t\|\mathbf{A} f\| P(\tau^{\mathbf{x}} \leq t),$$

where $\tau^{\mathbf{x}} := \inf\{t \geq 0 : |\mathbf{X}_t^{\mathbf{x}}| = R\}$. But $P(\tau^{\mathbf{x}} \leq t) \leq \sup\{P(\tau^{\mathbf{y}} \leq t) : |\mathbf{y}| = R+\varepsilon\}$, by the strong Markov property applied to the stopping time $\tau := \inf\{t \geq 0 : |\mathbf{X}_t^{\mathbf{x}}| = R + \varepsilon\}$. Now

$$P(\tau^{\mathbf{y}} \leq t) \leq P(\max_{0 \leq s \leq t} |\mathbf{X}_s^{\mathbf{y}} - \mathbf{y}| \geq \varepsilon) = \mathrm{o}(t) \qquad \text{as } t \downarrow 0,$$

uniformly for all $\mathbf{y}$ satisfying $|\mathbf{y}| = R + \varepsilon$. This argument extends to give (b) (see Exercise 6). To prove the local property (c) of $\hat{\mathbf{A}}$ consider $\mathbf{T}_t f(\mathbf{x}) - \mathbf{T}_t g(\mathbf{x}) = \mathrm{o}(t)$ as $t \downarrow 0$, by Exercise 8. ■

Example 1 (Brownian Motion) In the case $\boldsymbol{\mu} = 0, \boldsymbol{\sigma} = \mathbf{I}$, $\mathbf{X}_t^x = \mathbf{x} + \mathbf{B}_t, t \geq 0$ is k-dimensional standard Brownian motion starting at $\mathbf{x}$ and $\mathbf{A} = \frac{1}{2}\Delta$.

Proposition 15.3 (Initial-Value Problem) *Let $\boldsymbol{\mu}(\cdot), \boldsymbol{\sigma}(\cdot)$ be Lipschitzian. Adopt the notation of Definition 15.1.*

a. *If $f \in \mathcal{D}$, then $\mathbf{T}_t f \in \mathcal{D}$ for all $t \geq 0$, and $\hat{\mathbf{A}}\mathbf{T}_t f = \mathbf{T}_t\hat{\mathbf{A}} f$.*

b. *(Existence of a Solution) Let f satisfy the hypothesis* **1** *or* **2** *of Proposition 15.2. The function $u(t, \mathbf{x}) := \mathbf{T}_t f(\mathbf{x})$ satisfies Kolmogorov's backward equation*

$$\frac{\partial u(t, \mathbf{x})}{\partial t} = \hat{\mathbf{A}} u(t, \mathbf{x}) \qquad (t > 0, \mathbf{x} \in \mathbb{R}^k),$$

and the initial condition

$$u(t, \mathbf{x}) \to f(\mathbf{x}) \qquad \textit{as } t \downarrow 0, \ \textit{uniformly on } \mathbb{R}^k.$$

If, in addition, $u(t, \mathbf{x})$ is twice continuously differentiable in $\mathbf{x}$, for $t > 0$, then $\hat{\mathbf{A}} u(t, \mathbf{x}) = \mathbf{A} u(t, \mathbf{x})$.

c. *(Uniqueness) Suppose $v(t, \mathbf{x})$ is continuous on $[0, \infty) \times \mathbb{R}^k\}$, once continuously differentiable in $t > 0$ and twice continuously differentiable in $\mathbf{x} \in \mathbb{R}^k$, and satisfies*

$$\tfrac{\partial v(t,\mathbf{x})}{\partial t} = \mathbf{A} v(t, \mathbf{x}) \qquad (t > 0, \mathbf{x} \in \mathbb{R}^k),$$

$$\lim_{t \downarrow 0} v(t, \mathbf{x}) = f(\mathbf{x}) \qquad (\mathbf{x} \in \mathbb{R}^k),$$

where f is continuous and polynomially bounded. If $\mathbf{A} v(t, \mathbf{x})$ and $|\,\mathbf{grad}\, v(t, \mathbf{x})| = |(\partial/\partial x^{(1)}, \ldots, \partial/\partial x^{(k)}) v(t, \mathbf{x})|$ are polynomially bounded uniformly for every compact set of time points in $(0, \infty)$, then $v(t, \mathbf{x}) = u(t, \mathbf{x}) := \mathbb{E} f(\mathbf{X}_t^{\mathbf{x}})$.

Proof For (a),

$$\frac{1}{h}\left(\mathbf{T}_h(\mathbf{T}_t f) - \mathbf{T}_t f\right) = \frac{\mathbf{T}_t(\mathbf{T}_h f - f)}{h} \to \mathbf{T}_t \hat{\mathbf{A}} f$$

in "sup norm," since $\|\mathbf{T}_t g\| \leq \|g\|$ for all bounded measurable g. Part (b) follows from Proposition 15.2 and part **1**.

For part (c) on uniqueness, for each $t > 0$ use Itô's lemma to the function $w(s, \mathbf{x}) = v(t - s, \mathbf{x})$ to get

$$\mathbb{E}v(\varepsilon, \mathbf{X}^{\mathbf{x}}_{t-\varepsilon}) - v(t, \mathbf{x}) = \mathbb{E}\int_{\varepsilon}^{t} \left\{-v_0(r, \mathbf{X}^{\mathbf{x}}_{t-r}) + \mathbf{A}v(r, \mathbf{X}^{\mathbf{x}}_{t-r})\right\} dr = 0,$$

where $v_0(t, \mathbf{y}) := \frac{\partial}{\partial t} v(t, \mathbf{y})$. Let $\varepsilon \downarrow 0$, to get $\mathbb{E}v(0, \mathbf{X}^{\mathbf{x}}_t) = v(t, \mathbf{x})$, i.e., $v(t, \mathbf{x}) = \mathbb{E}f(\mathbf{X}^{\mathbf{x}}_t)$. ■

Proposition 15.4 (Dirichlet Problem) *Let G be a bounded open subset of $\mathbb{R}^k$. Assume that $\boldsymbol{\mu}(\cdot), \boldsymbol{\sigma}(\cdot)$ are Lipschitzian and that, for some i, $a_{ii}(\mathbf{x}) := |\boldsymbol{\sigma}_i(\mathbf{x})|^2 > 0$ for $\mathbf{x} \in \bar{G}$. Suppose v is a twice continuously differentiable function on G, continuous on $\bar{G}$, satisfying*

$$\begin{aligned} \mathbf{A}v(\mathbf{x}) &= -g(\mathbf{x}) \qquad (\mathbf{x} \in G), \\ v(\mathbf{x}) &= f(\mathbf{x}) \qquad (\mathbf{x} \in \partial G), \end{aligned}$$

where f and g are (given) continuous functions on ∂G and $\bar{G}$, respectively. Assume that v can be extended to a twice continuously differentiable function on $\mathbb{R}^k$. Then

$$v(\mathbf{x}) = \mathbb{E}f(\mathbf{X}^{\mathbf{x}}_\tau) + \mathbb{E}\int_0^\tau g(\mathbf{X}^{\mathbf{x}}_s)ds \qquad (\mathbf{x} \in \bar{G}),$$

where $\tau := \inf\{t \geq 0 : \mathbf{X}^{\mathbf{x}}_t \in \partial G\}$.

Proof Note that v can be taken to be twice continuously differentiable on $\mathbb{R}^k$ with compact support. Apply Itô's lemma to $\{v(\mathbf{X}^{\mathbf{x}}_t) : t \geq 0)\}$, and then use the optional stopping theorem. ■

Remark 15.2 It is known that twice-continuously differentiable functions on a bounded open set $G \subset \mathbb{R}^k$ with derivatives having continuous limits at the boundary ∂G have extensions to continuously differentiable functions on $\mathbb{R}^k$ if ∂G is C^1-smooth.[1]

15.1 Feynman–Kac̑ Formula for Multidimensional Diffusion

Proposition 15.5 (Feynman–Kac̑ Formula) *Let $\boldsymbol{\mu}(\cdot), \boldsymbol{\sigma}(\cdot)$ be Lipschitzian. Suppose $u(t, \mathbf{x})$ is a continuous function on $[0, \infty) \times \mathbb{R}^k$, once continuously differentiable in t for $t > 0$, and twice continuously differentiable in $\mathbf{x}$ on $\mathbb{R}^k$, satisfying*

---

[1] For example, see Stein (1970), Chapter VI.

$$\frac{\partial u(t,\mathbf{x})}{\partial t} = \mathbf{A}u(t,\mathbf{x}) + v(\mathbf{x})u(t,\mathbf{x}) \qquad (t>0, \mathbf{x}\in\mathbb{R}^k), \qquad u(0,\mathbf{x}) = f(\mathbf{x}), \tag{15.4}$$

where f is a polynomially bounded continuous function on $\mathbb{R}^k$, and v is a continuous function on $\mathbb{R}^k$ that is bounded above. If **grad** $u(t,\mathbf{x})$ *and* $\mathbf{A}u(t,\mathbf{x})$ *are polynomially bounded in* $\mathbf{x}$ *uniformly for every compact set of t in* $(0,\infty)$*, then*

$$u(t,\mathbf{x}) = \mathbb{E}(f(\mathbf{X}_t^{\mathbf{x}})\exp\{\int_0^t v(\mathbf{X}_s^{\mathbf{x}})ds\}) \qquad (t\ge 0, \mathbf{x}\in\mathbb{R}^k).$$

Proof For each $t>0$, apply Itô's lemma to the $(k+1)\times(k+1)$ stochastic differential equation for

$$\mathbf{Y}(s) := (\mathbf{Y}_1(s), \mathbf{Y}_2(s)) = (\mathbf{X}_s^{\mathbf{x}}, \int_0^s v(\mathbf{X}_{s'}^{\mathbf{x}})ds'),\ \varphi(s,\mathbf{y}) = u(t-s, y_1)\exp(y_2),$$

where $\mathbf{y} = (y_1, y_2)$. ■

Remark 15.3 An obviously equivalent form of (15.4) frequently occurs in applications with $v(\mathbf{x})u(t,\mathbf{x})$ replaced by $-v(\mathbf{x})u(t,\mathbf{x})$ under the condition[2] that v is bounded below, often, in fact, nonnegative. The Feynman–Kaĉ representation admits a particularly nice probabilistic interpretation in this context; see Exercise 3. Also, in the case that λ is a (positive) constant, note that the Feynman–Kaĉ formula may be obtained with the aid of Proposition 15.2 by transforming to an initial value problem for $w(t,\mathbf{x}) = e^{\lambda t}u(t,\mathbf{x})$ by introducing an integrating factor.

Example 2 (Integrated Geometric Brownian Motion) Let $B = \{B_t : t\ge 0\}$ denote standard Brownian motion started at 0. Also let $\mu, \sigma^2 > 0$ be given real number parameters and define

$$\mathcal{A}_t := \int_0^t \exp\{\mu s + \sigma B_s\}ds, \quad t\ge 0. \tag{15.5}$$

The stochastic process $\mathcal{A} = \{\mathcal{A}_t : t\ge 0\}$ arises in mathematical finance where one considers so-called *Asian options*. These are derivative contracts that are written to be contingent on the value of averaged price process $\mathcal{A}$ over some specified time horizon T, rather than over the geometric Brownian motion model $Z_t = \exp\{\mu t + \sigma B_t\}, 0\le t\le T$, often used in finance to model underlying prices in continuous time. One motivation for the development of such contracts is that they are less susceptible to short-term manipulation of the underlying asset.[3] The

[2] See Chen et al. (2003) for a derivation of the Itô (1965) condition on the potential for validity of a Feynman-Kac formula.

[3] Also see Thomann and Waymire (2003) for an alternative application to contracts written on natural resources.

stochastic process (15.5) is also of interest in the physics of *polymers* and other disordered systems.[4]

As an illustration of the use of the Feynman–Kaĉ formula, we will use it here to derive a second-order linear parabolic equation governing the density $a(t, y)$ of $\mathcal{A}_t, t \geq 0$. In particular, although $\mathcal{A}_t$ is clearly not a Markov process, we will derive the following surprising fact.[5]

Proposition 15.6 *The density defined by* $P(\mathcal{A}_t \in dy) = a(t, y)dy$ *exists for* $t > 0$ *and satisfies*

$$\frac{\partial a}{\partial t} = \frac{\partial^2}{\partial y^2}\{\frac{1}{2}\sigma^2 y^2 a\} - \frac{\partial}{\partial y}\{[(\frac{1}{2}\sigma^2 + \mu)y + 1]a\}, t > 0, a(0, y) = \delta_0(y).$$

Proof One may directly check that $\mathcal{A}_t$ has a density (Exercise 1). Let $Z_t := \mu t + \sigma B_t$ and define $X_t^x := e^{-Z_t}(x + \mathcal{A}_t), t \geq 0$. Then one may readily check that the jointly defined process $\{(X_t^x, Z_t) : t \geq 0\}$ is a Markov process with drift vector $\mathbf{m} = (1 + \frac{1}{2}\sigma^2 x - \mu x, \mu)$ and diffusion matrix $\mathbf{D} = \begin{pmatrix} \sigma^2 x^2 & -\sigma^2 x \\ -\sigma^2 x & \sigma^2 \end{pmatrix}$ (Exercise 2). Let $\hat{a}(t, \xi) := \mathbb{E}\exp(i\xi\mathcal{A}_t), t \geq 0$. Also define $\varphi(t, z, \xi) = \mathbb{E}\exp(i\xi e^z \mathcal{A}_t) = \mathbb{E}_{0,z} h_\xi(X_t, Z_t)$, where $h_\xi(x, z) = \exp(i\xi x e^\xi)$. Then an application of the Feynman–Kac formula to $\varphi(t, z, \xi)$ yields

$$\frac{\partial \varphi}{\partial t} = A^{(z)}\varphi + i\xi e^z \varphi, \quad \varphi(0, z, \xi) = 1, \tag{15.6}$$

where $A^{(z)} = \mu\frac{\partial}{\partial z} + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial z^2}$. Now observe that $\hat{a}(t, \xi) = \varphi(t, 0, \xi) = \mathbb{E}\exp(i\xi\mathcal{A}_t)$. Thus, performing the indicated differentiations in (15.6), it follows that

$$\frac{\partial \hat{a}}{\partial t} = \frac{1}{2}\sigma^2\xi^2\frac{\partial^2 \hat{a}}{\partial \xi^2} + (\frac{1}{2}\sigma^2 + \mu)\xi\frac{\partial \hat{a}}{\partial \xi} + i\xi e^z \hat{a}.$$

Now apply the Fourier inversion formula $a(t, y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\xi y}\hat{a}(t, \xi)d\xi$, to complete the proof. ∎

[4] See Comtet et al. (1998) and references therein to the mathematical physics literature.

[5] See Bhattacharya et al. (2001) and Rogers and Shi (1995) for this and for $v(t, x) := \mathbb{E}f(x + \mathcal{A}_t)$ for a class of homogeneous functions f. More recently, a relatively simple explicit calculation of the Laplace transform of the distribution of the hitting time has been found by Metzler (2013).

15.2 Kolmogorov Forward Equation (The Fokker–Planck Equation)

We close this chapter with Kolmogorov's forward equation, also referred to as the Fokker–Planck equation. Let us begin with the form of the backward and forward equations, respectively, in the context of the transition probability densities of an unrestricted multidimensional diffusion on $S = \mathbb{R}^k$.

$$\frac{\partial p}{\partial t} = \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}(\mathbf{x})\frac{\partial^2 p}{\partial x^{(i)}\partial x^{(j)}} + \sum_{i=1}^{k}\mu^{(i)}(\mathbf{x})\frac{\partial p}{\partial x^{(i)}},$$

$$\frac{\partial p}{\partial t} = \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\partial^2\left(a_{ij}(\mathbf{y})p\right)}{\partial y^{(i)}\partial y^{(j)}} - \sum_{i=1}^{k}\frac{\partial\left(\mu^{(i)}(\mathbf{y})p\right)}{\partial y^{(i)}}, \tag{15.7}$$

Consider a diffusion $\{X_t\}$ on $S = \mathbb{R}^k$ whose coefficients are Lipschitzian, with nonsingular diffusion coefficient. Let $p(t; x, y)$ be the transition probability density of $\{X_t\}$. Letting $f = \mathbf{1}_B$ the semigroup property may be expressed as

$$\int_B p(s+t; x, y)dy = \int_B (\int_S p(t; z, y)p(s; x, z)dz)dy \tag{15.8}$$

for all Borel sets $B \subset S$. This implies the *Chapman–Kolmogorov equation*

$$p(s+t; x, y) = \int_S p(t; z, y)p(s; x, z)dz = (\mathbf{T}_s f)(x), \tag{15.9}$$

where $f(z) := p(t; z, y)$.

A somewhat informal derivation of the backward equation for p may be based on (15.9) as follows. One has

$$(\mathbf{A}f)(x) = \lim_{s\downarrow 0}\frac{(T_s f)(x) - f(x)}{s}. \tag{15.10}$$

But the right side equals $\partial p(t; x, y)/\partial t$, and the left side is $\mathbf{A}p(t; x, y)$. This is precisely the desired *backward equation for* p.

In applications to physical sciences, the equation of greater interest is often Kolmogorov's forward equation, or the Fokker–Planck equation, governing the probability density function of X_t, when X_0 has an arbitrary initial distribution π. Suppose for simplicity that π has a density g. Then the density of X_t is given by

$$(\mathbf{T}_t^* g)(y) := \int_S g(x)p(t; x, y)dx. \tag{15.11}$$

The operator $\mathbf{T}_t^*$ transforms a probability density g into another probability density. More generally, it transforms any integrable g into an integrable function $\mathbf{T}_t^* g$. It is *adjoint* (*transpose*) to $\mathbf{T}_t$ in the sense that inserting a factor $f(y)$ and integrating (15.11) with respect to Lebesgue measure dy, one obtains

$$\langle T_t^* g, f\rangle := \int_S (T_t^* g)(y) f(y) dy = \int_S g(x)(\mathbf{T}_t f)(x) dx = \langle g, \mathbf{T}_t f\rangle. \tag{15.12}$$

Here $\langle u, v\rangle = \int u(x)v(x)\,dx$. If f is twice continuously differentiable and vanishes outside a compact subset of S, then one may differentiate with respect to t in (15.12) and interchange the orders of integration and differentiation to get, using the backward equation,

$$\begin{aligned}\langle \frac{\partial}{\partial t}\mathbf{T}_t^* g, f\rangle &= \langle g, \frac{\partial}{\partial t}\mathbf{T}_t f\rangle = \langle g, \mathbf{T}_t \mathbf{A} f\rangle \\ &= \langle T_t^* g, \mathbf{A} f\rangle \equiv \int_S (T_t^* g)(y)(\mathbf{A} f)(y) dy. \end{aligned}\tag{15.13}$$

Now, assuming that f, h are both twice continuously differentiable and that f vanishes outside a compact set, and h is integrable with respect to Lebesgue measure, integration by parts yields

$$\begin{aligned}\langle h, \mathbf{A} f\rangle &= \int_S h(y)[\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}(\mathbf{y})\frac{\partial^2 f}{\partial y^{(i)}\partial y^{(j)}} + \sum_{i=1}^{k}\mu^{(i)}(\mathbf{y})\frac{\partial f}{\partial y^{(i)}}]d\mathbf{y} \\ &= \int_S [\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\partial^2 (a_{ij}(\mathbf{y})h)}{\partial y^{(i)}\partial y^{(j)}} - \sum_{i=1}^{k}\frac{\partial(\mu^{(i)}(\mathbf{y})h)}{\partial y^{(i)}}] f(y) d\mathbf{y} \\ &= \langle \mathbf{A}^* h, f\rangle, \end{aligned}\tag{15.14}$$

where $\mathbf{A}^*$ is the *formal adjoint* of $\mathbf{A}$ defined by

$$(\mathbf{A}^* h)(y) = \mathrm{div}\{-h(y)\mu^{(i)}(y) + \frac{1}{2}\mathrm{grad}(\sum_j \frac{\partial}{\partial y_j})(a_{ij}(y)h(y))\}_{1\le i\le k}. \tag{15.15}$$

Applying (15.14) in (15.13), with $\mathbf{T}_t^* g$ in place of h, one gets

$$\langle \frac{\partial}{\partial t}\mathbf{T}_t^* g, f\rangle = \langle \mathbf{A}^*\mathbf{T}_t^* g, f\rangle. \tag{15.16}$$

Since (15.16) holds for sufficiently many functions f, all infinitely differentiable functions vanishing outside some closed, bounded subset of S for instance, we get (Exercise 11),

$$\frac{\partial}{\partial t}(\mathbf{T}_t^* g)(y) = \mathbf{A}^*(\mathbf{T}_t^* g)(y). \tag{15.17}$$

That is,

$$\int_S \frac{\partial p(t; x, y)}{\partial t} g(x)dx = \int_S [\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\partial^2(a_{ij}(\mathbf{y})p)}{\partial y^{(i)}\partial y^{(j)}} - \sum_{i=1}^{k}\frac{\partial(\mu^{(i)}(\mathbf{y})p)}{\partial y^{(i)}}]g(x)dx. \tag{15.18}$$

Since (15.18) holds for sufficiently many functions g, we get Kolmogorov's forward equation for the transition probability density p (Exercise 12).

Remark 15.4 (A Physical Description of the Forward Equation) Given an initial concentration $c_0(y)$ of solute in a fluid, the concentration $c(t, y)$ at y at time t is given by

$$c(t, y) = \int_{\mathbb{R}^3} c_0(z)p(t; z, y)dz, \tag{15.19}$$

where $p(t; z, y)$ is the transition probability density of the position process of an individual solute particle. Einstein's derivation of the diffusion equation for $A = \frac{1}{2}D\Delta$ is a special case (see Einstein 1905), where $p(t; z, y)$ is the transition probability density of a three-dimensional Brownian motion with dispersion matrix $D\mathbf{I}_3$, $D > 0$. Now, Kolmogorov's forward equation for $c(t, y)$ may be expressed as

$$\frac{\partial c(t, y)}{\partial t} = \mathbf{A}^* c(t, y) = -\mathrm{div} J(t, y), \tag{15.20}$$

with $J = (J_1, J_2, J_3)$ given by

$$J_i(t, y) = c(t, y)\mu^{(i)}(y) - \frac{1}{2}\sum_j \frac{\partial}{\partial y}(a_{ij}(y)c(t, y)), i = 1, 2, 3. \tag{15.21}$$

The forward equation is also referred to as the *Fokker–Planck equation* in this context. From a physical perspective, one may further understand the quantity J defined by (15.21) as follows. The increase in the amount of solute in a small region $[y, y + \Delta y]$ during a small time interval $[t, t + \Delta t]$ is approximately

$$\frac{\partial c(t, y)}{\partial t}\Delta t \Delta y. \tag{15.22}$$

On the other hand, if $v(t, y) = (v_1(t, y), v_2(t, y), v_3(t, y))$ denotes the velocity of the solute at y at time t, moving as a continuum, then $v(t, y)\Delta t$ flows into the region at y during $[t, t + \Delta t]$. Hence the amount of solute that flows into the region at y during $[t, t + \Delta t]$ is approximately $v(t, y)c(t, y)\Delta t$, while the amount passing out at $y + \Delta y$ during $[t, t + \Delta t]$ is approximately $v(t, y + \Delta y)c(t, y + \Delta y)\Delta t$. Therefore, the change in the amount of solute in $[y, y + \Delta y]$ during $[t, t + \Delta t]$ is approximately

$$[v(t,y)c(t,y) - v(t,y+\Delta y)c(t,y+\Delta y)]\Delta t. \tag{15.23}$$

Assuming mass cannot escape, equating (15.22) and (15.23) and dividing by $\Delta t \Delta y$, after letting Δt, $\Delta y = (\Delta y^{(1)}, \ldots, \Delta^{(k)})$ go to zero, one has

$$\frac{\partial}{\partial t} c(t,y) = -\mathrm{div}(v(t,y)c(t,y)). \tag{15.24}$$

The equation (15.24) is generally referred to as the *equation of continuity* or the *equation of mass conservation*. The quantity $J(t,y) = v(t,y)c(t,y)$ is called the *flux of the solute*, which is seen to be the *rate* at which the solute mass per unit volume is displaced. In the present case, therefore, the flux is given by (15.21). Also see Exercise 4 when there is an additional source, other than this flux, from which particles are created or annihilated at the rate $g(t,y)$ at y at time t.

Exercises

1. Show that $\mathcal{A}_t$ in Example 2 has a density for each $t > 0$.
2. Prove the Markov property for the process $\{(\mathcal{A}_t, Z_t) : t \geq 0\}$ in Example 2 and calculate the drift vector $\mathbf{m}$ and diffusion matrix $\mathbf{a}$. [*Hint*: Use Itô's lemma.]
3. Assume the conditions for the Feynman–Kaĉ formula stated in Proposition 15.5, with $v(\mathbf{x}) = -\lambda(\mathbf{x})$ where $\lambda(\mathbf{x})$ is a nonnegative continuous function. Enlarge the probability space $(\Omega, \mathcal{F}, P)$ for $\mathbf{X}^{\mathbf{x}}, \mathbf{x} \in \mathbb{R}^k$ to adjoin a (possibly infinite) nonnegative random variable ζ such that conditionally given $\sigma\{\mathbf{X}_s : 0 \leq s \leq t\}$, ζ is distributed with *hazard rate* $h(t) = \lambda(\mathbf{X}_t), t \geq 0$; that is, $P(\zeta > t | \sigma\{\mathbf{X}_s : 0 \leq s \leq t\}) := e^{-\int_0^t \lambda(\mathbf{X}_s^{\mathbf{x}})ds}, t \geq 0$. Let $\tilde{\mathbf{X}}_t^{\mathbf{x}} = \mathbf{X}_t^{\mathbf{x}}, 0 \leq t < \xi$. (i) Show that $u(t,\mathbf{x}) = \mathbb{E} f(\tilde{\mathbf{X}}_t^{\mathbf{x}}), t \geq 0$. (ii) Prove that $\{\tilde{X}_t : t \geq 0\}$ is a (nonconservative) Markov process with generator $\tilde{A} = A + v(\cdot) = A - \lambda(\cdot)$, where $(A\lambda(\cdot))f(x) = Af(x) - \lambda(x)f$. Consider the special case $\lambda(x) = \lambda > 0$ is a constant.
4. (Sources and Sinks: Duhamel's principle)

   (i) Consider the one-dimensional diffusion equation given by $\frac{\partial u}{\partial t} c(t,y) = -\frac{d}{dy}(\mu(y)c(t,y)) + \frac{1}{2}(\sigma^2(y)c(t,y)) + g(t,y), t > 0, c(0+,y) = c_0(y), y \in \mathbb{R}$, for Lipschitz coefficients $\mu(y), \sigma^2(y) > 0$. Use that $c_0(y)$ is integrable and $g(t,y)$ is a bounded, continuous function such that $\int_{\mathbb{R}} |g(t,x)| p(t;x,y)dx < \infty$, where $p(t;x,y)$ is transition probability density[6] of the diffusion with generator $A = \frac{1}{2}\sigma^2(x)\frac{d^2}{dx^2} + \mu(x)\frac{d}{dx}$.

[6] See Friedman (1964) for the existence of a density.

Check that $c(t,y) = \int_{\mathbb{R}} c_0(x)p(t;x,y)dx + \int_0^t \int_{\mathbb{R}} g(s,x)p(t-s;x,y)dxds$ is a solution, and provide a probability heuristic for the formula. State and solve the corresponding backward equation. [*Hint* :Assume $\lim_{s\downarrow 0}\int_{\mathbb{R}} g(t-s,x)p(t;x,y)dy = g(t,x)$.]

(ii) Consider the one-dimensional nonhomogeneous equation with Dirichlet boundary conditions: $\frac{\partial}{\partial t}u(t,x) = \mu(x)\frac{\partial u}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2 u}{\partial x^2} + h(t,x),\ t > 0, a < x < b,\ u(t,a+) = u(t,b-) = 0$, and $u(0+,x) = f(x), a \le x \le b$, where $\mu(x), \sigma^2(x) > 0$ are Lipschitz, h is a bounded, continuous function on $[0,\infty)\times[a,b]$, and f is a bounded, continuous function on $[a,b]$ such that $f(a) = f(b) = 0$. Check[7] that $u(t,x) = \int_0^t \int_{[a,b]} h(s,y)q(t-s,x,y)dyds + \int_{[a,b]} f(y)q(t,x,y)dy,\ t > 0$ is a solution, where $q(t,x,y)$ is the transition probability density component of a diffusion absorbed at a, b, with generator $A = \frac{1}{2}\sigma^2(x)\frac{d^2}{dx^2} + \mu(x)\frac{d}{dx}$. State and solve the corresponding Fokker–Planck equation.

(iii) State and prove analogs of (i), (ii) in multidimension.

5. Suppose that $g(\cdot)$ is a real-valued, bounded, nonanticipative functional on $[0,T]$. Prove that

$$t \mapsto \left(\int_0^t g(u)dB_u\right)^m g(t) \quad \text{belongs to} \quad \mathcal{M}[0,T]$$

for all positive integers m.

6. (i) Let m be a positive integer, and g a nonanticipative functional on $[0,T]$ such that $\mathbb{E}\int_0^T g^{2m}(t)dt < \infty$. Prove that

$$\mathbb{E}\left(\int_0^T g(t)dB_t\right)^{2m} \le (m(2m-1))^m T^{m-1}\int_0^T \mathbb{E}g^{2m}(t)dt.$$

[*Hint*: For bounded nonanticipative step functionals g, use Itô's lemma to get

$$J_t := \mathbb{E}\left(\int_0^t g(s)dB_s\right)^{2m} = m(2m-1)\int_0^t \mathbb{E}\{(\int_0^s g(u)dB_u)^{2m-2}g^2(s)\}ds,$$

so that J_t increases as t increases, and

$$\begin{aligned}\frac{dJ_t}{dt} &= m(2m-1)\mathbb{E}\{(\int_0^t g(u)dB_u)^{2m-2}g^2(t)\}\\ &\le m(2m-1)\{\mathbb{E}(\int_0^t g(u)dB_u)^{2m}\}^{(2m-2)/2m}(\mathbb{E}g^{2m}(t))^{1/m}\\ &= m(2m-1)J_t^{1-\frac{1}{m}}(\mathbb{E}g^{2m}(t))^{1/m},\end{aligned}$$

by Hölder's inequality.]

[7] See Friedman (1964) for the existence of a density.

(ii) Extend (i) to multidimension, i.e., for nonanticipative functionals $\mathbf{f} = \{(f_1(t), \ldots, f_k(t)) : 0 \le t \le T\}$ satisfying $\mathbb{E}\int_0^T f_i^{2m}(t)dt < \infty$ $(1 \le i \le k)$, prove that

$$\mathbb{E}(\int_0^T \mathbf{f}(t) \cdot d\mathbf{B}_t)^{2m} \le k^{2m-1}(m(2m-1))^m T^{m-1} \sum_{i=1}^k \mathbb{E}\int_0^T f_i^{2m}(t)dt.$$

7. (i) In addition to the hypothesis of Theorem 7.1 assume that $\mathbb{E}X_\alpha^{2m} < \infty$, where m is a positive integer. Prove that for $0 \le t - \alpha \le 1$,

$$\begin{aligned}\|X_t - X_\alpha\|_{2m} &\le c_1(m, M)[(t-\alpha)\|\mu(X_\alpha)\|_{2m} + (t-\alpha)^{1/2}\|\sigma(X_\alpha)\|_{2m}] \\ &= c_1(m, M)\varphi(t-\alpha),\end{aligned}$$

say, with $\varphi(t-\alpha) \downarrow 0$ as $t \downarrow \alpha$). Here $\|\cdot\|_{2m}$ is the L^{2m}-norm for random variables, and $c_1(m, M)$ depends only on m and M. [*Hint*: Let $\delta_t = \mathbb{E}\max_{\alpha\le s\le t}|X_s - X_\alpha|^{2m} = \mathbb{E}\max_{\alpha\le s\le t}|\int_\alpha^s(\mu(X_u) - \mu(X_\alpha)|du + (s-\alpha)\mu(X_\alpha) + \int_\alpha^s(\sigma(X_u) - \sigma(X_\alpha))dB_u + \sigma(X_\alpha)(B_s - B_\alpha)|^{2m} \le 4^{2m-1}[\cdots]$, and then use Grownwall's inequality.]

(ii) Deduce from (i) that $\sup_{\alpha\le s\le t}\mathbb{E}X_s^{2m} < \infty$ for all $t \ge \alpha$.

(iii) Extend (i), (ii) to multidimension.

8. Let $\boldsymbol{\mu}(\cdot), \boldsymbol{\sigma}(\cdot)$ be Lipschitzian. Prove that (a) for every $\varepsilon > 0$,

$$P(\max_{0\le s\le t}|\mathbf{X}_s^{\mathbf{x}} - \mathbf{x}| \ge \varepsilon) = \mathrm{o}(t) \qquad \text{as } t \downarrow 0,$$

uniformly for every bounded set of $\mathbf{x}$'s. Show that the convergence is uniform for all $\mathbf{x} \in \mathbb{R}^k$ if $\boldsymbol{\mu}(\cdot), \boldsymbol{\sigma}(\cdot)$ are bounded as well as Lipschitzian.[*Hint*:

$$\begin{aligned}&P(\max_{0\le s\le t}|\mathbf{X}_s^{\mathbf{x}} - \mathbf{x}| \ge \varepsilon) \\ &\le P(\int_0^t|\boldsymbol{\mu}(\mathbf{X}_s^{\mathbf{x}})|ds \ge \frac{\varepsilon}{2}) + P(\max_{0\le s\le t}|\int_0^s \boldsymbol{\sigma}(\mathbf{X}_u^{\mathbf{x}})d\mathbf{B}_u| \ge \frac{\varepsilon}{2}) = I_1 + I_2,\end{aligned}$$

say. To estimate I_1 note that $|\boldsymbol{\mu}(\mathbf{X}_s^{\mathbf{x}})| \le |\boldsymbol{\mu}(\mathbf{x})| + M|\mathbf{X}_s^{\mathbf{x}} - \mathbf{x}|$, and use Chebyshev's inequality for the fourth moment and Exercise 7(iii). To estimate I_2 use Doob's L^p-maximal inequality (15.3) for the $p = 4$th moment, Exercise 6(ii) with $m = 2$, and Exercise 7(iii).] (b) Prove that I_1, I_2 go to zero as $t \downarrow 0$ uniformly on $\mathbb{R}^k$ if $\mu(\cdot)$ and $\sigma(\cdot)$ are also bounded.

9. Show that if (15.16) holds for all infinitely differentiable functions that vanish outside a compact set, then the forward equations hold.

10. Let $\mu(\cdot), \sigma(\cdot)$ be Lipschitzian. Show that the solution $X_{t,\varepsilon}^x$ of $X_{t,\varepsilon}^x = x + \int_0^t \mu(X_{s,\varepsilon}^x)ds + \int_0^t \varepsilon\sigma(X_{s,\varepsilon}^x)dB_s$ converges as $\varepsilon \to 0$, to the solution of the ordinary differential equation $\frac{dY_t^x}{dt} = \mu(Y_t^x)$, $Y_0^x = x$, in the norm distance $||X^x - Y^x||$ defined by $||Z||^2 = \mathbb{E}\max_{0\le s\le t}|Z_s|^2$.

11. Supply the details to derive (15.17).
12. Show that the class of infinitely differentiable functions g vanishing outside compact sets is sufficient to establish (15.18).

Chapter 16
Probabilistic Solution of the Classical Dirichlet Problem

The classical Dirichlet problem seeks to find functions with specified boundary values on a domain for which the function is harmonic on the interior. The intrinsic connection between the Laplacian operator and Brownian motion leads to a beautiful interplay between pde and stochastic calculus presented in this chapter.

In physics, a conservative force is the gradient of a potential, which is a scalar function. The Laplacian of this function is the divergence of its gradient, and its vanishing implies equilibrium in many problems in physics, such as gravitation, electricity, and magnetism. The Dirichlet problem is to determine the potential in a (usually bounded) region given its values on the boundary. It also arises in the problem of equilibrium temperature distribution.

Throughout this chapter, we consider standard Brownian motions $\mathbf{B}^x = \{\mathbf{B}_t^x : t \geq 0\}$ in $\mathbb{R}^k$ (dimension $k \geq 2$), starting at $\mathbf{x} \in \mathbb{R}^k$.

A real-valued function f defined on a nonempty open subset G of $\mathbb{R}^k$ is said to be *harmonic* in G if it is twice differentiable in G and satisfies

$$\Delta f(x) = 0 \qquad \text{for all } \mathbf{x} = (x_1, \ldots, x_k) \in G \tag{16.1}$$

where Δ is the *Laplacean*

$$\Delta = \sum_{i=1}^{k} \partial^2/\partial x_i^2. \tag{16.2}$$

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_16

Suppose now that G is bounded. The *classical Dirichlet problem* consists in finding, for each continuous real-valued function φ on ∂G, a function u on $\overline{G} = G \cup \partial G$ that is (i) harmonic in G and (ii) continuous on $\overline{G}$ with $u(x) = \varphi(x)$ $\forall x \in \partial G$. It is known that the problem is not always solvable in this generality (see Itô and McKean 1963, pp. 261-264, for the famous counterexample known as *Lebesgue's thorn*), and that certain smoothness assumptions need to be imposed on ∂G to ensure the existence of a u satisfying (i) and (ii). One such condition is that every point x on the boundary is a Poincaré point.

Definition 16.1 A point $\mathbf{x} \in \partial G$ is said to be a *Poincaré point* if there exists a truncated cone C with vertex at $\mathbf{x}$, $C\backslash\{\mathbf{x}\} \subset \overline{G}^c$.

Let $C([0, \infty) \to \mathbb{R}^k)$ denote the set of all continuous functions on $[0, \infty)$ into $\mathbb{R}^k$. Define

$$\overline{\tau}(f) = \overline{\tau}_{\partial G}(f) := \inf\{t \geq 0 : f(t) \in \partial G\}, \quad f \in C([0, \infty) \to \mathbb{R}^k). \tag{16.3}$$

Our main objective in this chapter is to prove the theorem stated below of Kakutani and Doob. This will be followed by an application to calculate the probability $u(\mathbf{x})$ that a k-dimensional Brownian motion starting at $\mathbf{x} \in G := \{\mathbf{y} : R_1 < |\mathbf{y} - \mathbf{a}| < R_2\}$, where $\mathbf{a} \in \mathbb{R}^k, 0 < R_1 < R_2$, will reach $F_1 = \{\mathbf{y} \in \mathbb{R}^k : |\mathbf{y} - \mathbf{a}| = R_1\}$ before $F_2 = \{\mathbf{y} \in \mathbb{R}^k : |\mathbf{y} - \mathbf{a}| = R_2\}$. Observe that $u = \varphi$ on $\partial G = F_1 \cup F_2$, where $\varphi(\mathbf{x}) = 1, \mathbf{x} \in F_1$ and $\varphi(\mathbf{x}) = 0, \mathbf{x} \in F_2$. Moreover

$$u(\mathbf{x}) = \mathbb{E}\varphi(B^{\mathbf{x}}_{\overline{\tau}(B^x)}) \quad \mathbf{x} \in \overline{G}, \tag{16.4}$$

In view of the next theorem, this probability may be analytically computed by solving an appropriate Dirichlet problem.

Theorem 16.1 (Doob-Kakutani) *Let G be a bounded nonempty open subset of $\mathbb{R}^k$*

a. *For every real-valued bounded Borel measurable function φ on ∂G (regarded as a metric space with the Euclidean distance), the function*

$$u(\mathbf{x}) := \mathbb{E}\varphi(B^{\mathbf{x}}_{\overline{\tau}(B^x)}) \quad \mathbf{x} \in \overline{G}, \tag{16.5}$$

is harmonic in G.

b. *Suppose that every point of ∂G is a Poincaré point. Then, for each bounded continuous φ on ∂G, u is the unique solution of the Dirichlet problem satisfying (i) and (ii).*

We will prove the theorem in several steps, each step stated as a lemma.

Lemma 1 *Under the hypothesis of part (a) of the theorem, u has the mean value property: For every $r > 0$ such that $B(\mathbf{x} : r) \equiv \{\mathbf{y} : |\mathbf{y} - \mathbf{x}| < r\} \subseteq G$ one has*

$$u(\mathbf{x}) = \int_{S^{k-1}} u(\mathbf{x} + r\boldsymbol{\theta})\mu(d\boldsymbol{\theta}) \tag{16.6}$$

where μ is the uniform distribution (normalized surface area measure) on the unit sphere $S^{k-1} = \{\mathbf{y} : |\mathbf{y} - \mathbf{x}| = 1\}$.

Proof Fix $x \in \partial G$, and let $r > 0$ be such that $B(\mathbf{x} : r) \subseteq G$. Let $\eta := \inf\{t \geq 0 : |B_t^x - x| = r\}$. Since the after–$\eta$ process $B_\eta^{x+} \equiv \{B_{\eta+t}^x : t \geq 0\}$ hits ∂G at the same point as B^x does, one has the following relations in view of the strong Markov property with respect to the stopping time η: Conditioning on the pre-η σ-field $\mathcal{G}_\eta$, one has

$$\begin{aligned}
\varphi(B^x_{\overline{\tau}(B^x)}) &= \varphi((B_\eta^{x+})_{\overline{\tau}(B_\eta^{x+})}) \\
u(x) &= \mathbb{E}(\mathbb{E}[\varphi((B_\eta^{x+})_{\overline{\tau}(B_\eta^{x+})})|\mathcal{G}_\eta]) \\
&= \mathbb{E}(\mathbb{E}[\varphi((B^y)_{\overline{\tau}(B^y)})]_{y=B_\eta^x}) \\
&= \mathbb{E}u(B_\eta^x).
\end{aligned} \tag{16.7}$$

Since the distribution of B^x is invariant under orthogonal transformations around x, i.e., $B^x - x$ is distributed as B^0, the same as that of $O(B^x - x)$, for every orthogonal transformation O, it follows that the distribution of B_η^x is the uniform distribution on the sphere $\{y : |y - x| = r\}$. By a change of variables $y \to r\theta$, or $\theta = y/r$, $\theta \in S^{k-1}$, one arrives at (16.6) from (16.7). ■

Lemma 2 *If a function u has the mean value property in G, then it is infinitely differentiable and harmonic in G.*

Proof Let $a \in G$. Find $R > 0$ such that $\overline{B(a; 2R)} \subseteq G$. We first show that u is infinitely differentiable in $B(a : R)$. For this, choose an infinitely differentiable radial function $\psi(x) = g(|x|)$ whose support is contained in $B(0 : R)$ and such that $\int \psi(x)dx = 1$ (for example, let $g(r) = c_k \exp\{-(\frac{R^2}{4} - r^2)^{-1}\}$ for $0 \leq r < R/3$, $g(r) = 0$ for $r \geq \frac{R}{2}$, such that $c_k \int_0^R g(r)r^{k-1}dr = 1$). Define

$$\overline{u}(x) = \begin{cases} u(x) & \text{for } x \in \overline{B(a : 2R)} \\ 0 & \text{for } x \in \mathbb{R}^k \backslash \overline{B(a : 2R)}. \end{cases} \tag{16.8}$$

For $y \in B(a : R)$ one has, using the polar transformation,

$$\begin{aligned}
(\overline{u} \star \psi)(y) \equiv \int \overline{u}(y + x)\psi(x)dx &= \int_{B(0:R)} u(y + x)g(|x|)dx \\
&= c_k \int_{S^{k-1}} \int_0^R u(y + r\theta)g(r)r^{k-1}dr\mu(d\theta)
\end{aligned}$$

$$
\begin{aligned}
&= c_k \int_0^R g(r) r^{k-1} \Big(\int_{S^{k-1}} u(y + r\theta)\mu(d\theta)\Big) dr \\
&= u(y),
\end{aligned}
\tag{16.9}
$$

where c_k is the surface area of S^{k-1}. Since $\overline{u} \star \psi$ is infinitely differentiable everywhere, u is infinitely differentiable in $B(a : R)$, by (16.9). Thus u is infinitely differentiable in G. To show that u is harmonic, again let $a \in G$. In a sufficiently small neighborhood of a (contained in G), Taylor expansion yields up to $O(|x-a|^3)$

$$
u(x) = u(a) + \sum_{i=1}^k (x_i - a_i)\Big(\frac{\partial u(x)}{\partial x_i}\Big)_a + \frac{1}{2}\sum_{i,j=1}^k (x_i - a_i)(x_j - a_j)\Big(\frac{\partial^2 u(x)}{\partial x_i \partial x_j}\Big)_a. \tag{16.10}
$$

Integrating both sides with respect to the uniform distribution $\mu_p(dx)$, say, on the sphere $\{x : |x - a| = \rho\} = S(a : \rho)$, for sufficiently small ρ, and using the mean value property of u, one gets

$$
u(a) = u(a) + \frac{\rho^2}{2k}\sum_{i=1}^k \Big(\frac{\partial^2 u(x)}{\partial x_i^2}\Big)_a + O(\rho^3). \tag{16.11}
$$

Note that (16.11) results from rotational invariance of μ_ρ. Specifically,

$$
\begin{aligned}
&(1) \quad \int_{S(a:\rho)} (x_i - a_i)\mu_\rho(dx) = 0 \qquad \forall i, \\
&(2) \quad \int_{S(a:\rho)} (x_i - a_i)(x_j - a_j)\mu_\rho(dx) = 0 \qquad \forall\, i \neq j, \\
&(3) \quad \int_{S(a:\rho)} (x_i - a_i)^2\mu_\rho(dx) = \int_{S(a:\rho)} (x_1 - a_1)^2\mu_\rho(dx) \\
&\qquad = \int_{S(a:\rho)} \frac{1}{k}\sum_{i=1}^k (x_i - a_i)^2 \mu_\rho(dx) \qquad \forall\, i.
\end{aligned}
\tag{16.12}
$$

For (1) the μ_ρ–distribution of $(x_i - a_i)$ is the same as that of $-(x_i - a_i)$, (2) the μ_ρ–distribution of $x - a$ is the same as that of $x' - a'$, where $x'_i - a'_i = -(x_i - a_i)$ and $x'_j - a'_j = x_j - a_j$ $\forall\, j \neq i$, (3) the μ_ρ–distribution of $x_i - a_i$ is the same for all i. Now (16.11) implies

$$
\frac{\rho^2}{2k}\,\Delta u(a) = O(\rho^3) \qquad \text{as } \rho \downarrow 0, \tag{16.13}
$$

which is only possible if $\Delta u(a) = 0$. ■

Remark 16.1 This follows from a special case of Proposition 15.4 applied to balls, and Lemma 1(Exercise). But here is a direct proof. The converse of Lemma 1 is also true: *If u is a twice continuously differentiable function in G which is harmonic, then u has the mean value property.* To see this fix $a \in G$, and consider the function

$$f(r) = \int_{S^{k-1}} u(a + r\theta)\mu(d\theta) \qquad (r > 0 \text{ sufficiently small}).$$

Then

$$f'(r) = \int_{S^{k-1}} \Big(\frac{d}{dr} u(a + r\theta)\Big)\mu(d\theta) = \int_{S^{k-1}} \langle \nu, \mathbf{grad}\, u\rangle(a + r\theta)\mu(d\theta),$$

where $\langle\ ,\ \rangle$ denotes the Euclidean inner product and ν is the outward unit normal to S^{k-1} at θ, i.e., $\nu = \theta = (\theta_1, \theta_2, \ldots, \theta_k)$. Hence, by the Gauss divergence theorem,

$$\begin{aligned} f'(r) &= c_k' r^k \int_{B(0:1)} \nabla(\mathbf{grad}\, u)(a + rx)dx \\ &= c_k' r^k \int_{B(0:1)} (\Delta u)(a + rx)dx = 0, \end{aligned}$$

where c_k' is a normalizing constant ($c_k' \int_{B(0:1)} dx = 1$). Thus $r \to f(r)$ is a constant and, hence, $f(r) = u(a) \equiv \lim_{r \downarrow 0} f(r)$.

Lemmas 1 and 2 complete the proof of part (a) of Theorem 16.1. For part (b) we need to show that u is continuous on $\overline{G}$, i.e., $u(x) \equiv \mathbb{E}\varphi(B^x_{\overline{\tau}(B^x)}) \to u(b) = \varphi(b)$ as $x \to b$ for every $b \in \partial G$. To require this for every bounded continuous φ on ∂G is equivalent to the requirement that

$$B^x_{\overline{\tau}(B^x)} \text{ converges in distribution to } \delta_{\{b\}} \text{as} x \to b \quad (\forall\ b \in \partial G), \tag{16.14}$$

where (see (16.3) $B^x_{\overline{\tau}(B^x)} = \inf\{t \geq 0 : B^x_t \in \partial G\}$.

Lemma 3 *Let $b \in \partial G$. Suppose that for every $\delta > 0$, one has*

$$\lim_{\substack{x \to b, \\ x \in G}} P(\overline{\tau}(B^x) > \delta) = 0. \tag{16.15}$$

Then (16.14) holds.

Proof Let φ be a bounded continuous function on ∂G. Given $\varepsilon > 0$, let $\eta_1 > 0$ be such that $|\varphi(y) - \varphi(b)| < \varepsilon/3$ if $|y - b| < 2\eta_1$, $y \in \partial G$. Write $\|\varphi\|_\infty :=$ $\max\{|\varphi(y)| : y \in \partial G\}$. Then

$$
\begin{aligned}
|u(x) - u(b)| &= |\mathbb{E}\varphi(B^x_{\overline{\tau}(B^x)}) - \varphi(b)| \\
&\leq \mathbb{E}(|\varphi(B^x_{\overline{\tau}(B^x)}) - \varphi(b)| \cdot \mathbf{1}_{\{\overline{\tau}(B^x) \leq \delta\}}) + 2\|\varphi\|_\infty P(\overline{\tau}(B^x) > \delta) \\
&< \frac{\varepsilon}{3} P(\max_{0 \leq t \leq \delta} |B^x_t - b| < 2\eta_1, \overline{\tau}(B^x) \leq \delta) \\
&\quad + 2\|\varphi\|_\infty P(\max_{0 \leq t \leq \delta} |B^x_t - b| \geq 2\eta_1) + 2\|\varphi\|_\infty P(\overline{\tau}(B^x) > \delta). \quad (16.16)
\end{aligned}
$$

For $|x - b| < \eta_1$,

$$
\begin{aligned}
P(\max_{0 \leq t \leq \delta} |B^x_t - b| \geq 2\eta_1) &\leq P(\max_{0 \leq t \leq \delta} |B^x_t - x| \geq \eta_1) \\
&= P(\max_{0 \leq t \leq \delta} |B^0_t| \geq \eta_1) \longrightarrow 0 \quad \text{as } \delta \downarrow 0. \quad (16.17)
\end{aligned}
$$

Therefore, there exists $\delta_0 > 0$ such that, if $|x - b| < \eta_1$ then

$$
2\|\varphi\|_\infty P(\max_{0 \leq t \leq \delta_0} |B^x_t - b| \geq 2\eta_1) < \frac{\varepsilon}{3}. \quad (16.18)
$$

Also, in view of (16.15), one may find $\eta_2 > 0$ such that if $|x - b| < \eta_2$ then

$$
2\|\varphi\|_\infty P(\overline{\tau}(B^x) > \delta_0) < \frac{\varepsilon}{3}. \quad (16.19)
$$

From (16.16)–(16.19), one gets $|u(x) - u(b)| < \varepsilon$ if $|x - b| < \eta := \min\{\eta_1, \eta_2\}$. Thus, as $x \to b$, $u(x) \to u(b)$. ∎

Definition 16.2 A boundary point b is said to be *regular* if (16.15) holds.

We have shown that the regularity of all boundary points is *sufficient* for the well-posedness of the Dirichlet problem in G. We will later see that it is also *necessary*. However, a more amenable criterion is

$$
P(\tilde{\tau} \circ B^x = 0) = 1, \quad (16.20)
$$

where $\tilde{\tau}(f)$ is the *first time after zero* that f is in $\overline{G}^c$

$$
\tilde{\tau}(f) := \inf\{t > 0 : f(t) \in \overline{G}^c\} \qquad (f \in C([0, \infty) \to \mathbb{R}^k)).
$$

Note that if $x \in G$, then $\tilde{\tau} \circ B^x = \overline{\tau}(B^x) > 0$. Also note that $\overline{\tau} \circ B^x$ is a stopping time with respect to the filtration $\mathcal{F}_{t+}(= \cap_{\varepsilon > 0}\mathcal{F}_{t+\varepsilon})$, $t \geq 0$ (see BCPT[1] Exercise 12, p.73). Here $\mathcal{F}_t = \sigma(B_s : s \leq t)$. It is simple to show that a $\mathcal{F}_t$-Brownian motion

[1] Throughout, BCPT refers to Bhattacharya and Waymire (2016), A Basic Course in Probability Theory.

is als a $\{\mathcal{F}_{t+}\}$-Brownian motion (see Bhattacharya and Waymire 2021, Proposition 1.6, p. 79). It is not difficult to show that a $\{\mathcal{F}_t\}$-diffuson is also a $\{\mathcal{F}_{t+}\}$-diffusion (Exercise 4).

Lemma 4 *If (16.20) holds for a boundary point b, then it is a regular boundary point.*

Proof For each $\delta > 0$,

$$P(\overline{\tau}(B^x) > \delta) = P(\tilde{\tau} \circ B^x > \delta) \qquad \forall\, x \in G. \tag{16.21}$$

Thus, it is enough to show that $P(\tilde{\tau} \circ B^x > \delta) \to 0$ as $x \to b$ $(x \in G)$. Write $A_h = \{f \in C([0, \infty) : \mathbb{R}^k) : f(t) \in \overline{G} \text{ for } h \leq t \leq \delta\}$, $0 < h < \delta$. Then $[\tilde{\tau} > \delta] = \lim_{h \downarrow 0} A_h$, so that

$$P(\tilde{\tau} \circ B^x > \delta) = \lim_{h \downarrow 0} P(B^x \in A_h). \tag{16.22}$$

Now, by the Markov property, writing $\varphi(y) = P(B_t^y \in \overline{G} \text{ for } 0 \leq t \leq \delta - h)$, one has for $0 < h < \delta$,

$$\begin{aligned} P(B^x \in A_h) &= \mathbb{E}[(P(B_t^y \in \overline{G} \;\text{ for }\; 0 \leq t \leq \delta - h))_{y = B_h^x}] \\ &= \mathbb{E}\varphi(B_h^x) = \int_{\mathbb{R}^k} \varphi(y) p(h; x, y) dy. \end{aligned} \tag{16.23}$$

Since φ is bounded and $x \to p(h; x, y)$ is continuous, $x \to P(B^x \in A_h)$ is continuous on $\mathbb{R}^k$. Since (16.20) holds, $P(\tilde{\tau} \circ B^b > \delta) = 0 \;\forall\, \delta > 0$. Let $\varepsilon > 0$. Fix any $\delta > 0$. Then one may, by (16.22), find $h > 0$ such that

$$P(B^b \in A_h) < \varepsilon.$$

Now find $\eta > 0$ such that $P(B^x \in A_h) < \varepsilon$ if $|x - b| < \eta$, using the continuity of $x \to P(B^x \in A_h)$. Then

$$P(\tilde{\tau} \circ B^x > \delta) \leq P(B^x \in A_h) < \varepsilon.$$

■

Lemma 5 *If b is a Poincaré point of ∂G, and φ a continuous function on ∂G, then u is continuous at b.*

Proof By Blumenthal's zero–one law[2]

$$P(\tilde{\tau} \circ B^b = 0) = 0 \;\text{ or }\; 1, \tag{16.24}$$

[2] See BCPT, Proposition 11.6, p.198.

since $[\tilde{\tau} \circ B^b = 0] \in \mathcal{F}_{0+} = \cap_{t>0}\mathcal{F}_t$, where $\mathcal{F}_t = \sigma(X_s^b : s \le t\}$..

Let $\widetilde{C}$ be a cone with vertex at b and $C\backslash\{b\} = C_b$, say, a truncation of $\widetilde{C}$ contained in $\overline{G}^c$. By rotating $\widetilde{C}$ around b, one may find a finite number, say n, of rotations of $\widetilde{C}$ whose union is $\mathbb{R}^k$. If $P(\tilde{\tau} \circ B^b = 0) = 0$, then $P(\tilde{\tau} \circ B^b > 0) = 1$ so that one can find $\delta_0 > 0$ for which $P(\tilde{\tau} \circ B^b > \delta_0) > 1 - \frac{1}{2n}$, which in turn implies $P(B_t^b \in C_b^c$ for $0 < t \le \delta_0) > 1 - \frac{1}{2n}$. Now the probability of $[B_t^b \in C]$ is the same as that of $[B_t^b \in C_i]$ $(i = 1, \dots, n)$, where C_i is obtained from C by one of the n rotations of $\widetilde{C}$ whose union is $\mathbb{R}^k$. But $\cup_{i=1}^n C_i$ contains a ball $B(b : r)$, $r > 0$. Hence, one would have $P(B_t^b \in B(b : r)) < n\frac{1}{2n}$ for all $t \in (0, \delta_0]$, which is of course not true, since letting $t \downarrow 0$ $(t > 0)$ one would get $1 = \lim_{t\downarrow 0} P(B_t^b \in B(b : r)) < \frac{1}{2}$. ■

It only remains to prove *uniqueness*. This is a consequence of

Lemma 6 (The Maximum Principle) *Suppose G is a connected open set and u has the mean value property in G. Then u cannot attain its supremum or infimum in G, unless u is constant in G.*

Proof Suppose u attains its supremum at a point $x_0 \in G$. By the mean value property, for all sufficiently small p,

$$u(x_0) = \int_{S_p(x_0)} u(y)\mu_\rho(dy),$$

where μ_ρ is the normalized surface area measure on $S_\rho(x_0) = \{y : |y - x_0| = \rho\}$ (i.e., $\mu_\rho(dy)$ is the distribution of $y = x_0 + \rho\theta$, when θ has the rotation invariant distribution on S^{k-1}). Since u is continuous on $S_\rho(x_0)$, and μ_ρ assigns positive measure to every nonempty, open subset of S_ρ (i.e., *support* of μ_ρ is $S_\rho(x_0)$), and $u(y) \le u(x_0)$ on $S_\rho(x_0)$, it follows that $u(y) = u(x_0)$ $\forall$ $y \in S_\rho(x_0)$ (for all sufficiently small ρ). This means that the set $D \equiv \{x \in G : u(x) = u(x_0)\}$ is open, but D is clearly closed. Thus $G = D \cup (G\backslash D)$ is a union of two disjoint open sets. Since G is connected, this implies $D = G$. ■

Note that one can apply the argument to each connected component of an arbitrary open set G and conclude that *if $u \equiv 0$ on ∂G, then $u \equiv 0$* in G *(provided u has the mean value property in G and u is continuous on $\overline{G} = G \cup \partial G$*). The proof of the Theorem is now complete. If u_1, u_2 are two solutions, then $u_1 - u_2 = 0$ on ∂G and has the mean value property. If $u_1(x) > u_2(x)$ at some $x \in G$, $u_1(x) - u_2(x) > 0$, contradicting the fact that the maximum is on ∂G. ■

Example 1 (Application to Transience and Recurrence) Although Proposition 16.2 follows from results in Chapter 11, we give a proof based on Theorem 16.1, with some new computations. As an application, we consider the problem of transience and recurrence of k-dimensional standard Brownian motion.

Proposition 16.2 *Two-dimensional Brownian motion is recurrent in the sense that given any* $\mathbf{a} \in \mathbb{R}^2$ *and* $\varepsilon > 0$*,*

$$P(\eta^{\mathbf{x}}_{B(\mathbf{a},\varepsilon)} = \infty) = 1, \quad \forall\, \mathbf{x} \in \mathbb{R}^2, \tag{16.25}$$

where

$$\eta^{\mathbf{x}}_{B(\mathbf{a},\varepsilon)} := \sup\{t \geq 0 : |B^{\mathbf{x}}_t - \mathbf{a}| \leq \varepsilon\} \tag{16.26}$$

is the time of the last visit to $B(\mathbf{a}, \varepsilon)$*. Every higher dimensional Brownian motion on* $\mathbb{R}^k$*,* $k \geq 3$*, is transient in the sense that*

$$P\left(|B^{\mathbf{x}}_t| \longrightarrow \infty \;\; as \;\; t \to \infty\right) = 1 \qquad \forall\, \mathbf{x} \in \mathbb{R}^k. \tag{16.27}$$

Proof Let $0 < R_1 < R_2$. In Theorem 16.1 take

$$G = \{\mathbf{x} : R_1 < |\mathbf{x} - \mathbf{a}| < R_2\}.$$

Then

$$\begin{aligned} \partial G &= \{\mathbf{x} : |\mathbf{x} - \mathbf{a}| - R_1\} \cup \{\mathbf{x} : |\mathbf{x} - \mathbf{a}| = R_2\}, \\ \tau &= \overline{\tau}_{\partial G}(B^x) \end{aligned}$$

Define

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{for } |\mathbf{x} - \mathbf{a}| = R_1, \\ 0 & \text{for } |\mathbf{x} - \mathbf{a}| = R_2. \end{cases}$$

Then starting from $\mathbf{x} \in G$, the probability $u(\mathbf{x})$ of reaching the inner sphere of radius R_1 before reaching the outer sphere of radius R_2 may be expressed as $u(\mathbf{x}) = \mathbb{E}\varphi(B^{\mathbf{x}}_\tau)$, satisfies

$$\begin{aligned} \Delta u(\mathbf{x}) &= 0 \qquad && \forall\, \mathbf{x} \in G, \\ u(\mathbf{x}) &= \varphi(\mathbf{x}) \qquad && \forall\, \mathbf{x} \in \partial G. \end{aligned} \tag{16.28}$$

For simplicity, let $\mathbf{a} = 0$ first. Then the function (16.28) is invariant under rotations so that $u(\mathbf{x}) = g(|\mathbf{x}|)$ for some g which is continuous on $[R_1, R_2]$ and twice (actually, infinitely) continuously differentiable on (R_1, R_2). Applying the Laplacian to $g(|\mathbf{x}|)$, one gets

$$\Delta u(\mathbf{x}) = g''(|\mathbf{x}|) + \frac{k-1}{|\mathbf{x}|}\, g'(|\mathbf{x}|). \tag{16.29}$$

In other words, g satisfies

$$\begin{aligned} g''(r) + \frac{k-1}{r}\, g'(r) = 0 \qquad & R_1 < r < R_2, \\ g(R_1) = 1, \qquad & g(R_2) = 0. \end{aligned} \tag{16.30}$$

Writing $f = g'$, one gets $f'(r) + \frac{k-1}{r} f(r) = 0$, whose general solution is

$$f(r) = c_1 (\frac{R_1}{r})^{k-1}. \tag{16.31}$$

Hence $g(r) = c_1 \int_{R_1}^{r} (R_1/r')^{k-1} dr' + c_2$, i.e.,

$$g(r) = \begin{cases} c_1 R_1 \log \frac{r}{R_1} + c_2 & \text{if } k = 2 \\ -c_1 \frac{R_1^{k-1}}{k-2} (\frac{1}{r^{k-2}} - \frac{1}{R_1^{k-2}}) + c_2 & \text{if } k > 2, \end{cases}$$

where c_1, c_2 are determined by the boundary conditions in (16.30):

$$g(r) = \begin{cases} \frac{\log R_2 - \log r}{\log R_2 - \log R_1} & \text{if } k = 2 \\ \frac{\left(\frac{R_1}{r}\right)^{k-2} - \left(\frac{R_1}{R_2}\right)^{k-2}}{1 - \left(\frac{R_1}{R_2}\right)^{k-2}} & \text{if } k > 2. \end{cases} \tag{16.32}$$

Substituting $r = |\mathbf{x}|$ in (16.32), one gets $u(\mathbf{x})$. For arbitrary $\mathbf{a}$ use translation invariance, of Brownian motion to obtain, writing $\tau_{R_i}^{\mathbf{x}}$ for $\tau_{\partial B(\mathbf{a}:R_i)} \circ B^{\mathbf{x}}$ $(i = 1, 2)$,

$$P\left(\tau_{R_1}^{\mathbf{x}} < \tau_{R_2}^{\mathbf{x}}\right) = \begin{cases} \dfrac{\log R_2 - \log |\mathbf{x} - \mathbf{a}|}{\log R_2 - \log R_1} & \text{if } k = 2 \\ \dfrac{(R_1/|\mathbf{x} - \mathbf{a}|)^{k-2} - (R_1/R_2)^{k-2}}{1 - (R_1/R_2)^{k-2}} & \text{if } k > 2, \end{cases} \tag{16.33}$$

for $R_1 \le |\mathbf{x} - \mathbf{a}| \le R_2$. Letting $R_2 \uparrow \infty$ in (16.33), one arrives at

$$P\left(\tau_{R_1}^{\mathbf{x}} < \infty\right) = \begin{cases} 1 & \text{for } |\mathbf{x} - \mathbf{a}| \ge R_1, \text{if } k = 2, \\ (\frac{R_1}{|\mathbf{x}-\mathbf{a}|})^{k-2} & \text{for } |\mathbf{x} - \mathbf{a}| \ge R_1 \text{if} k > 2. \end{cases} \tag{16.34}$$

The completion of the proof is left as (Exercise 3). ■

In (16.33) if one lets $R_1 \downarrow 0$, then one gets ($\forall\ k \ge 2$), writing $\tau_{\{\mathbf{a}\}}^{\mathbf{x}} = \inf\{t \ge 0 : B_t^{\mathbf{x}} = \mathbf{a}\}$,

$$P(\tau_{\{\mathbf{a}\}}^{\mathbf{x}} < \tau_{R_2}^{\mathbf{x}}) = 0 \qquad \forall\ \mathbf{x} \ne \mathbf{a},\ |\mathbf{x} - \mathbf{a}| \le R_2,$$

so that, $\forall\ k \ge 2$,

$$P(\tau_{\{\mathbf{a}\}}^{\mathbf{x}} < \infty) = 0 \qquad \forall\ \mathbf{x} \ne \mathbf{a}. \tag{16.35}$$

Thus, the probability of ever hitting any given point (other than the starting point) is zero. This is in contrast with the one-dimensional case.

Example 2 (Poisson Equation) Before we close this subsection, we briefly consider *Poisson's equation:* For a given bounded continuous differentiable f on G find a v, which is continuous on $\overline{G}$, and satisfies

$$\begin{aligned} \Delta v(x) &= -f(x) \qquad x \in G, \\ v(x) &= 0 \qquad x \in \partial G, \end{aligned} \tag{16.36}$$

where G is a bounded open set with a C^1-smooth boundary. This is a special case of Proposition 15.4, but we present a proof based on the results of this chapter and some new computations Assume that ∂G is smooth enough and f is such that f and its first derivatives may be continuously extended to all of $\mathbb{R}^k$. Let this extension $\overline{f}$ be so chosen as to have compact support (for example, if G is a sphere or ellipsoid and the first order derivatives of f are uniformly continuous such an extension is easily shown to be possible). Assume $k > 2$. The function

$$w(x) = \frac{1}{c_k} \int_{\mathbb{R}^k} \frac{\overline{f}(y)}{|x-y|} \, dy \tag{16.37}$$

satisfies the equation

$$\Delta w(x) = -\overline{f}(x) \qquad x \in \mathbb{R}^k. \tag{16.38}$$

(Exercise 2; here c_k is the surface area of S^{k-1}). The solution of (16.36) is then *equivalent to the solution of the Dirichlet problem*

$$\begin{aligned} \Delta h(x) &= 0 \qquad \forall\, x \in G, \\ h(x) &= w(x) \qquad \forall\, x \in \partial G. \end{aligned} \tag{16.39}$$

For if h solves (16.39), then $v(x) \equiv w(x) - h(x) (x \in \overline{G})$ solves (16.36). Conversely, if $v(x)$ solves (16.36), then $h(x) \equiv w(x) - v(x)$ solves (16.39).

Exercises

1. In the proof of Lemma 5, it is said that $\mathbb{R}^k$ is the union of a finite number of rotations of the cone $\widetilde{C}$ (around b). Prove this. [*Hint*:It is enough to prove that the sphere $S(b : 1) := \{x : |x - b| = 1\}$ is the union of a finite number of intersections of the sphere with rotated cones. For this use compactness of $S(b : 1)$.]
2. Prove (16.38).
3. Use (16.34) to complete the proof of Proposition 16.2.
4. Consider a diffusion on $\mathbb{R}^k$ (or on a domain $D \subset \mathbb{R}^k$). Let $\mathcal{F}_t := \sigma(X_s : 0 \le s \le t)$. Show that $\{X_s : s \ge 0\}$ is a diffusion with respect to the filtration $\mathcal{F}_{t+}$. [*Hint*:The conditional distribution of X_t, given $\mathcal{F}_{s+\varepsilon}$, has the density $p(t - s - \varepsilon; x, y)$ on $[X_{s+\varepsilon} = x]$. Let $\varepsilon \downarrow 0$.]

Chapter 17
The Functional Central Limit Theorem for Ergodic Markov Processes

In this chapter a broadly applicable functional central limit theorem for continuous parameter ergodic Markov processes is developed. This uses the martingale central limit theorem applied to functions in the range of the infinitesimal generator. This chapter includes an example application to dispersion in a periodic media and another to obtain the seminal formula of Taylor and Aris for long-time dispersion of solutes embedded in fluid flow.

Consider an arbitrary Markov process X with stationary transition probabilities $p(t; x, dy)$ on a metric space S. A semigroup of transition operators T_t, $t \geq 0$, can be defined on the space $\mathbb{B}(S)$ of real-valued, bounded (Borel) measurable functions on S by

$$T_t f(x) = \mathbb{E}^x f(X_t) \equiv \int_S f(y) p(t; x, dy), \quad x \in S, f \in \mathbb{B}(S). \tag{17.1}$$

(Also see Chapter 2 on semigroups.) The *semigroup* property $T_{t+s} = T_t T_s = T_s T_t$ is a direct consequence of the Markov property. Moreover, if $\mathbb{B}(S)$ is given the norm $||f|| = \sup_{x \in S} |f(x)|$, then the averaging defining $T_t f(x)$ implies the *contraction* property $||T_t f|| \leq ||f||$. Moreover T_t is a *positive* operator in the sense that $f \geq 0$ pointwise on S implies that $T_t f \geq 0$ pointwise on S for each $t \geq 0$. The *infinitesimal generator* $(A, \mathcal{D}_A)$ of the Markov process X, or its associated semigroup, with respect to this norm is defined for $f \in \mathcal{D}_A$ by

$$Af = g \quad \text{if } \lim_{t \downarrow 0} ||\frac{T_t f - f}{t} - g|| = 0, \tag{17.2}$$

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_17

for some $g \in \mathbb{B}(S)$. The set $\mathcal{D}_A$, so-defined, is called the *domain* of A.

Suppose that the Markov process X has a unique invariant probability π. Then the Hilbert space $L^2(S, \pi)$ contains $\mathbb{B}(S)$. Moreover, using the invariance of π, one can check that $T_t, t \geq 0$ is a contraction semigroup of positive linear operators for the L^2-norm $||\cdot||_2$ on $L^2(S, \pi)$, as well. Moreover, an infinitesimal generator $\hat{A}$ can be defined on a domain $\mathcal{D}_{\hat{A}} \subseteq L^2(S, \pi)$ in the same limit but with the norm replaced by $||\cdot||_2$. Note that $\mathcal{D}_A \subseteq \mathcal{D}_{\hat{A}}$. Hence the generator $\hat{A}$ is an *extension* of A. Having said all of this, we will continue to use $(A, \mathcal{D}_A)$ for the generator on $L^2(S, \pi)$. In general the choice of the function space depends on what one is trying to quantify. Here is a repeat result from Chapter 2.

Lemma 1 *For $f \in \mathcal{D}_A$, (i) $T_t f \in \mathcal{D}_A$ for all $t > 0$, and (ii) $AT_t f = T_t Af$ for all $t > 0$.*

Proof By definition, and using commutativity and contraction properties of $T_t, t \geq 0$, one has

$$
\begin{aligned}
||(T_s(T_t f) - T_t f)/s - T_t Af|| &= ||(T_t\{(T_s f) - f)/s - Af\}|| \\
&\leq ||(T_s f - f)/s - Af|| \to 0 \quad \text{as } s \downarrow 0.
\end{aligned}
$$

■

Lemma 2 (Dynkin's Martingale) *Suppose that $X = \{X_t : t \geq 0\}$ is a Markov process on a metric space S, having right-continuous sample paths. Assume that for $f \in \mathcal{D}_A$, $(s, \omega) \to Af X_s(\omega)$ is progressively measurable (see BCPT[1] pp. 59–60). Then*

$$
Z_t := f(X_t) - \int_0^t Af(X_s)ds, \; t \geq 0, \tag{17.3}
$$

is a martingale with respect to $\mathcal{F}_t = \sigma(X_s : 0 \leq s \leq t), t \geq 0$.

Proof Note that the conditional distribution of the after-s process X_s^+ given $\mathcal{F}_s$ is the same as the distribution of X, starting at X_s. Also, using $T_s Af = AT_s f$, one has

$$
\begin{aligned}
\mathbb{E}(Z_{t+s}|\mathcal{F}_s) &= \mathbb{E}(f(X_{t+s})|\mathcal{F}_s) - \int_0^s Af(X_u)du - \mathbb{E}(\int_s^{s+t} Af(X_u)du|\mathcal{F}_s) \\
&= T_t f(X_s) - \int_0^s Af(X_u)du - \int_0^t T_u Af(X_s)du \\
&= T_t f(X_s) - \int_0^t AT_u f(X_s)du - \int_0^s Af(X_u)du.
\end{aligned} \tag{17.4}
$$

[1] Throughout, BCPT refers to Bhattacharya and Waymire (2016), A Basic Course in Probability Theory.

The middle integral in the last expression equals $\int_0^t \frac{d}{du} T_u f(X_s) du$ which, by the fundamental theorem of calculus (see Theorem 2.7), is $T_t f(X_s) - f(X_s)$. The martingale property of Z now follows. ■

Remark 17.1 The assumptions[2] of progressive-measurability implies integrability of the integrand almost surely defining Z_t. This will continue to be assumed without further mention. One may note that in applications to diffusions X, the sample paths are continuous functions, making this a nonissue. We will, in general, assume that the right-continuous Markov process is ergodic, and that S is a Polish space equipped with its Borel σ-field.

Remark 17.2 For diffusions defined on all of $\mathbb{R}^k$, Itô's lemma guarantees the martingale property of Z. However, the above Lemma 2 applies to general continuous parameter ergodic Markov processes.

Theorem 17.1 (Functional Central Limit Theorem for Ergodic Markov Processes[3]) *Suppose that X is a stationary, ergodic Markov process on S with transition probability $p(t; x, dy)$ and invariant probability π. Let $(A, \mathcal{D}_A)$ denote the generator on $L^2(S, \pi)$. Assume that f belongs to the range $\mathcal{R}_A$ of $(A, \mathcal{D}_A)$ with $Ag = f$ for some $g \in \mathcal{D}_A$. Then, letting $B = \{B_t : t \geq 0\}$ denote a standard Brownian motion*

a. $\int_S f(x)\pi(dx) = 0$.
b. $\{\frac{1}{\sqrt{n}} \int_0^{nt} f(X_s) ds : t \geq 0\}_n \Rightarrow \{\sigma B_t : t \geq 0\}$ *as* $n \to \infty$.
c. $\sigma^2 = -2\langle f, g\rangle_\pi = -2 \int_S f(x)g(x)\pi(dx)$.

Proof For (a) observe that $Ag = \frac{d}{dt} T_t g|_{t=0}$, so that by the invariance of π, one has

$$\int_S f d\pi = \int_S Ag d\pi = \frac{d}{dt} \int_S T_t g d\pi|_{t=0} = \frac{d}{dt} \int_S g d\pi = 0.$$

The proof of part (b) will follow from a version of the martingale functional central limit theorem originally due to Billingsley (1961) and Ibragimov (1963) (see Bhattacharya and Waymire (2022), Theorem 15.5). One has

$$T_t g(x) = g(x) + \int_0^t T_s Ag(x) ds \quad \pi - \text{a.e.}, \tag{17.5}$$

so that, by Lemma 2,

$$Y_n(g) := g(X_n) - \int_0^n Ag(X_s) ds, \quad n = 1, 2, \ldots, \tag{17.6}$$

[2] See BCPT, Proposition 3.6.

[3] Bhattacharya (1982).

is a square integrable martingale, and the difference sequence

$$\Delta_n(g) = g(X_{n+1}) - g(X_n) - \int_n^{n+1} Ag(X_s)ds, \quad n = 0, 1, 2 \ldots, \tag{17.7}$$

is stationary and ergodic. In particular, therefore,

$$\mathbb{E}\Delta_n(g) = 0, \quad n = 1, 2, \ldots. \tag{17.8}$$

So one has by the martingale functional central limit theorem that

$$Z_n(t) = n^{-1\overline{2}}(Y_{[nt]}(g) + (nt - [nt])\Delta_{[nt]}(g)), \quad t \geq 0, \tag{17.9}$$

converges in distribution to σB with

$$\sigma^2 = \mathbb{E}(\Delta_0(g))^2 = \mathbb{E}\{g(X_1) - g(X_0) - \int_0^1 Ag(X_s)ds\}^2. \tag{17.10}$$

Now,

$$\begin{aligned}
&|Z_n(t) + n^{-\frac{1}{2}} \int_0^{nt} Ag(X_s)ds| \\
&\quad \leq n^{-\frac{1}{2}}(|g(X_{[nt]})| + |\Delta_{[nt]}(g)| + \int_{[nt]}^{[nt]+1} |Ag(X_s)|ds) \\
&\quad = n^{-\frac{1}{2}}(I_1([nt]) + I_2([nt]) + I_3([nt]), \text{ say.}
\end{aligned} \tag{17.11}$$

Let us check that each of $n^{-\frac{1}{2}} I_j([nt])$, $j = 1, 2, 3$ is almost surely $o(1)$ as $n \to \infty$ by a simple Borel-Cantelli argument. Namely, for each $j = 1, 2, 3$, one has for any $\varepsilon > 0$, that

$$\begin{aligned}
\sum_{m=1}^{\infty} P(I_j(m) > \varepsilon\sqrt{m}) &= \sum_{m=1}^{\infty} P(I_j(1) > \varepsilon\sqrt{m}) \\
&\leq \int_0^{\infty} P(I_j(1) > \varepsilon\sqrt{s})ds \\
&= \frac{2}{\varepsilon^2} \int_0^{\infty} P(I_j(1) > r)rdr = \mathbb{E}I_j^2(1) < \infty.
\end{aligned} \tag{17.12}$$

Thus, for any $t_0 > 0$, one has P_π-a.s. that

$$\sup_{0 \leq t \leq t_0} n^{-\frac{1}{2}}(I_1([nt]) + I_2([nt]) + I_3([nt]) \to 0, \quad \text{a.s. as } n \to \infty.$$

Finally to compute σ in terms of the given parameters of the process, observe that, using the orthogonality of the martingale differences and stationarity,

$$\begin{aligned}
&\sigma^2 \\
&= \mathbb{E}[g(X_1) - g(X_0) - \int_0^1 f(X_s)ds]^2 \\
&= n[\mathbb{E}(g(X_{\frac{1}{n}}) - g(X_0))^2 \\
&\quad + \mathbb{E}(\int_0^{\frac{1}{n}} Ag(X_s)ds)^2 - 2\mathbb{E}(g(X_{\frac{1}{n}}) - g(X_0)) \int_0^{\frac{1}{n}} Ag(X_s)ds)].
\end{aligned}$$

Now,

$$\begin{aligned}
\mathbb{E}[g(X_{\frac{1}{n}}) - g(X_0)]^2 &= 2\int_S g^2 d\pi - 2\mathbb{E}g(X_0)g(X_{\frac{1}{n}}) \\
&= 2\int_S g^2 d\pi - 2\int_S gT_{\frac{1}{n}} g d\pi \\
&= 2\int_S g^2 d\pi - 2\langle g, T_{\frac{1}{n}} g\rangle_\pi \\
&= 2\int_S g^2 d\pi - 2\langle g, g + \frac{1}{n}Ag + o(\frac{1}{n})\rangle_\pi \\
&= -\frac{2}{n}\langle g, Ag\rangle_\pi + o(\frac{1}{n}).
\end{aligned}$$

Next

$$\mathbb{E}(\int_0^{\frac{1}{n}} Ag(X_s)ds)^2 \le \mathbb{E}\frac{1}{n}\int_0^{\frac{1}{n}} (Ag(X_s))^2 ds = \frac{1}{n^2}||Ag||^2.$$

Using these last two estimates together with the Cauchy-Schwarz inequality, one has $\mathbb{E}\big(g(X_{\frac{1}{n}}) - g(X_0)\big) \int_0^{\frac{1}{n}} Ag(X_s)ds\big) = O(n^{-\frac{3}{2}})$. Thus, one has $\sigma^2 = -2\langle g, Ag\rangle_\pi + o(1)$ as $n \to \infty$. Since σ^2 is independent of n, the proof is complete. ■

Remark 17.3 Note that we have used the notion of ergodicity of a discrete parameter sequence. This is considered in detail in Bhattacharya and Waymire (2022, Chapter 4). The notion is entirely analogous for continuous parameter processes.

There are a number of important questions that naturally arise with regard to the application of Theorem 17.1. Here is a list, which will be answered in the order that they occur below:

Q1 How large is the range space $\mathcal{R}_A$?
Q2 How can one guarantee that the asymptotic variance σ^2 is positive?
Q3 Under what additional conditions can one ensure that the asymptotic distribution holds regardless of the initial distribution for the Markov process?

Proposition 17.2 *In addition to the hypothesis of Theorem 17.1, assume that the semigroup $\{T_t\}$ is strongly continuous in $L^2(S,\pi)$, i.e., $\overline{\mathcal{D}}_A = L^2$. Then, (a) $\mathcal{R}_A$ is dense in $1^\perp$-the subspace of mean zero functions in L^2. (b) If one also assumes that* 0 *is an isolated point of the spectrum of* A*, then $\mathcal{R}_A = 1^\perp$.*

Proof (a) Let $\mathcal{N}_{\hat{A}}$ denote the null space of $\hat{A}$, i.e.,

$$\mathcal{N}_{\hat{A}} := \{f \in L^2 : \hat{A}f = 0\}.$$

Since $\hat{A}$ has a closed graph, $\mathcal{N}_{\hat{A}}$ is a closed subspace. We will now see that $\mathcal{N}_{\hat{A}}$ is a one-dimensional space of constants. For this, let $f \in \mathcal{N}_{\hat{A}}$. Applying Dynkin's martingale formula (Lemma 2), $\{f(X_t) : t \geq 0\}$ is a square integrable stationary martingale converging to some shift-invariant square integrable random variable Z (Theorem 12.3 in Bhattacharya and Waymire (2021).) Shift invariance and ergodicity imply that Z is a constant $\pi-$almost surely. Therefore,

$$f(X_t) = \mathbb{E}(Z|\sigma(X_s, 0 \leq s \leq t)) = Z, \text{ a.s..}$$

Thus $\mathcal{N}_{\hat{A}}$ is the one-dimensional subspace of constants, the eigenspace of 0. Next, note the general facts (i) the adjoint A^* is well-defined since $\mathcal{D}_A$ is dense in $L^2(S,\pi)$, and (ii) Range$(A) = \mathcal{N}_{A^*}^\perp$ (Exercise 1).[4] For (a), it is enough to show that $\mathcal{N}_{A^*} \supset \mathcal{N}_{\hat{A}}^\perp$, i.e., $\mathcal{N}_{A^*} \subset \mathcal{N}_{\hat{A}}$. For this, let $h \in \mathcal{N}_{\hat{A}^*}$. Then for all $g \in \mathcal{D}_{\hat{A}}$, $\langle \frac{d}{dt}T_t g, h\rangle = \langle \hat{A}T_t g, h\rangle = \langle T_t g, \hat{A}^* h\rangle = 0$, so that $\langle T_t g, h\rangle = \langle g, T_t^* h\rangle$ is constant in t. Hence $T_t^* h = h$ for all t. Now, $||T_t h - h||^2 = ||T_t h||^2 + ||h||^2 - 2\langle T_t h, h\rangle = 2||h||^2 - 2\langle T_t h, h\rangle = 2||h||^2 - 2\langle h, T_t^* h\rangle = 2||h||^2 - 2\langle h, h\rangle = 0$. This proves that $T_t h = h$ and, differentiating with respect to t, one gets $\hat{A}h = 0$, i.e., $h \in \mathcal{N}_{\hat{A}}$. For part (b), 0 belongs to the resolvent set $\rho(\hat{A})$, and for any $g \in 1^\perp$, $f := \int_0^\infty T_t g dt$ belongs to $\mathcal{D}_{\hat{A}}$, and (Exercise 3)

$$\hat{A}f = -g. \tag{17.13}$$

■

Remark 17.4 It may be shown that (b) is both necessary and sufficient for $\mathcal{R}_{\hat{A}} = 1^\perp$ (see Bhattacharya (1982)).

We next turn to question Q2. The following simple example indicates the necessity of addressing this question.

[4] See Yosida (1964), p.205.

Example 1 (Newman[5]*)* Let $t \to X_t, t \geq 0$ denote the deterministic periodic motion on the unit circle S^1. Then the uniform distribution π on the circle is the unique invariant probability. This strictly stationary Markov process is ergodic. If f is bounded and $\int_{S^1} f d\pi = 0$, then $\int_0^1 f(X_s)ds$ is bounded, and hence $\sigma^2 = 0$.

Example 2 Suppose that $\hat{A}$ is self-adjoint. Then under the hypothesis of Theorem 17.1, $\sigma^2 > 0$ for every nonzero $f \in \mathcal{R}_{\hat{A}}$. To see this, note that $\sigma^2 = -2\langle \hat{A}g, g\rangle$, where $\hat{A}g = f$. The spectral theorem[6] implies that $\sigma^2 > 0$.

For the remainder of this chapter, it will be convenient to assume that the semigroup is $\{T_t : t \geq 0\}$ is *strongly continuous* in $L^2(S, \pi)$, i.e., the domain $\mathcal{D}_A$ of the generator A is dense in $L^2(S, \pi)$.

Remark 17.5 Suppose π is an invariant probability for a transition probability $p(t; x, dy)$ and A the infinitesimal generator on $L^2(S, \pi)$. Since A is a closed operator, its null space

$$\mathcal{N}_A = \{h \in L^2(S, \pi) : Ah = 0\}$$

is closed. Therefore, given $f \in \mathcal{R}_A$, the range of A, there exists a unique $g \in \mathcal{D}_A \cap \mathcal{N}_A^{\perp}$ such that $Ag = f$. One may use this g in Theorem 17.1.

From the first part of the proof of part (a) of Proposition 17.2, the following result is an easy corollary.

Proposition 17.3 *Suppose that X is a stationary, Markov process on S with transition probability $p(t; x, dy)$ and invariant probability π. Then X is ergodic if and only if the null space is a one-dimensional space of constants. Equivalently, X is ergodic if and only if* 0 *is a simple eigenvalue of A.*

Our next task is to show that, at least under some reasonable additional conditions, the variance parameter σ^2 in Theorem 17.1(c) is strictly positive unless $f = 0$ π-a.s..

Proposition 17.4 *In addition to the hypothesis in Theorem 17.1, asssume that for each pair $(t, x) \in (0, \infty) \times S$, $p(t; x, dy)$ and $\pi(dy)$ are mutually absolutely continuous. Then if $f \in \mathcal{R}_{\hat{A}}$ and f is bounded π-a.s., then the variance parameter σ^2 in Theorem 17.1(c) is strictly positive, unless $f = 0$, π-a.s..*

Proof Suppose $f \in \mathcal{R}_{\hat{A}}$, f bounded, and $f \neq 0$. Let $g \in \mathcal{D}_{\hat{A}}$ be such that $\hat{A}g = f$. If possible suppose $\sigma^2 = 0$. Since differences of the martingale $Y_t = g(X_t) - \int_0^t f(X_s)ds$ over successive non-overlapping time intervals of equal length form a stationary sequence of martingale differences, each martingale difference must be zero a.s., so that one must have, choosing a separable version of the martingale if

[5] Personal communication by C.M. Newman.

[6] See Bhattacharya and Waymire (2022).

necessary:

$$Q(g(X_t) - g(X_0) - \int_0^t f(X_s)ds = 0 \text{ for all } t \geq 0) = 1,$$

where $(\Omega, \mathcal{F}, Q)$ is the underlying probability space for the Markov process $\{X_s : s \geq 0\}$. Thus, with probability one, $\int_0^t f(X_s)ds = g(X_t) - g(X_0)$ for all t. From this, one shows that[7] $g(X_t) - g(X_0) = 0$ Q-a.s.. Choose and fix $t > 0$. Since g is not an a.s. constant, there exist numbers $a < b$ and sets $A, B \in \mathcal{B}(S)$ such that

(i) $g < a$ on A, $g > b$ on B, and
(ii) $\pi(A) > 0,\ \pi(B) > 0.$

Hence, writing $p(t; x, y)$ for a strictly positive version of the density of $p(t; x, dy)$ with respect to $\pi(dy)$,

$$Q(g(X_0) < a, g(X_t) > b) = \int_{\{y:g(y)>b\}} \int_{\{x:g(x)<a\}} p(t; x, y)m(dx)\pi(dy) > 0,$$

which contradicts $g(X_t) - g(X_0) = 0$ a.s.. ■

Remark 17.6 Note that it is enough to assume the absolute continuity in Proposition 17.4 for sufficiently large t.

Let us now consider a progressively measurable Markov process $\{X_t : t > 0\}$ with state space S, initial distribution μ, and transition probability p defined on some probability space $(\Omega, \mathcal{A}, Q)$. As μ varies let the probability space vary. In this notation, $Q = Q_m$. The tail σ-field is $\mathcal{T} = \cap_{t \geq 0}[X_s : s \geq t]$. The tail σ-field is Q_μ-trivial if $Q_\mu(A) = 0$ or 1 for all $A \in \mathcal{T}$. The following proposition is essentially proved in Orey (1971), Proposition 4.3, pp. 19–20.

Proposition 17.5 *Let p be a transition probability admitting an invariant initial distribution μ. Then the following statements are equivalent:*

(a) *The tail σ-field is Q_μ-trivial for every initial distribution μ.*
(b) $||p(t; x, dy) - \pi(dy)||_v \to 0$ *as $t \to \infty$,for every $x \in S$.*

Here $||\nu||_v$ denotes the variation norm of a signed measure ν.

Remark 17.7 Note that if condition (b) (or (a)) of Proposition 17.5 holds, then for every π-integrable f one has

$$Q_\mu(\lim_{T\to\infty} T^{-1} \int_0^T f(X_s)ds = \int_S f dm) = 1, \tag{17.14}$$

[7] See Bhattacharya (1982).

whatever be the initial distribution μ. For the event C within parentheses is the same as $C \circ \theta_t = \{\lim_{T\to\infty} T^{-1} \int_t^{T+t} f(X_s)ds = \int_S f dm\}$ for every $t > 0$. Hence the left side of (17.14) equals $\mathbb{E}_\mu d(X_t)$ where $d(x) = Q_{\delta_x}(C)$ and $\mathbb{E}_\mu$ denotes expectation with respect to μ. Now $\mathbb{E}_m d(X_t) = Q_m(C) = 1$, by the ergodic theorem. Therefore,

$$\begin{aligned}
|Q_\mu(C) - 1| &= |\mathbb{E}_\mu d(X_t) - \mathbb{E}_\pi d(X_t)| \\
&= |\int_S \int_S d(y)(p(t; x, dy) - \pi(dy))\mu(dx)| \\
&\leq ||p(t; x, dy) - \pi(dy)||_v \to 0 \text{ as } t \to \infty.
\end{aligned}$$

Hence (17.14) holds.

We can now prove the following useful result.

Theorem 17.6 *Let p be a transition probability admitting an invariant initial distribution π. Assume that $||p(t; x, dy) - \pi(dy)||_v \to 0$ as $t \to \infty$, for all $x \in S$. Suppose $f \in \mathcal{R}_{\hat{A}}$. Then for every μ, the Q_μ-distributions of the random functions $n^{-\frac{1}{2}} \int_0^{nt} f(X_s)ds (t \geq 0)$, converge weakly to the Wiener measure with zero drift and variance parameter σ^2 given by Theorem 17.1(c).*

Proof Fix a probability measure μ on S. Write Z_n for the random function $n^{-\frac{1}{2}} \int_0^{nt} f(X_s)ds (t \geq 0)$. Let ψ be a real valued bounded continuous function on $C[0, \infty)$. For $h > 0$ let $Z_{n,h}$ denote the random function $n^{-\frac{1}{2}} \int_h^{nt+h} f(X_s)ds (t \geq 0)$. Then $\mathbb{E}_\mu \psi(Z_{n,h}) = \mathbb{E}_{\mu_h} \psi(Z_n)$, where μ_h is the Q_μ-distribution of X_h. Note that

$$|\mathbb{E}_{\mu_h} \psi(Z_n) - \mathbb{E}_m \psi(Z_n)| \leq ||\psi||_\infty \int_S ||p(h; x, dy) - m(dy)||_v \mu(dx) \to 0$$

as $h \to \infty$, uniformly for all n. Choose $h(n) \to \infty$, $h(n) = o(n^{-\frac{1}{2}})$. Then by Theorem 17.1, $\mathbb{E}_m \psi(Z_n) \to \int_{C[0,\infty)} \psi dW_{\sigma^2}$, where W_{σ^2} is the limiting Wiener measure. One has, therefore, $\lim_n \mathbb{E}_\mu \psi(Z_{n,h}) = \int_{C[0,\infty)} \psi dW_{\sigma^2}$. By the ergodic theorem and (17.14),

$$\sup_{t\geq 0} |Z_{n,h(n)}(t) - n^{-\frac{1}{2}} \int_0^{h(n)+nt} f(X_s)ds| \leq n^{-\frac{1}{2}} \int_0^{h(n)} |f(X_s)|ds \to 0$$

in Q_μ-probability as $n \to \infty$. Finally, the change of time

$$\theta_n(t) = (t - h(n)/n) \vee 0,$$

applied to the process $n^{-\frac{1}{2}} \int_0^{h(n)+nt} f(X_s)ds$, shows as $n \to \infty$, (see Billingsley (1968, pp. 144–145)),

$$\mathbb{E}_\mu \psi(Z_n) \to \int_{C[0,\infty)} \psi dW_{\sigma^2}.$$

■

17.1 A Functional Central Limit Theorem for Diffusions with Periodic Coefficients

For this example we will assume a symmetric positive definite matrix $\mathbf{a}(x) = ((a_{ij}(x)))_{1 \le i,j \le k}$ for each $x \in \mathbb{R}^k$, a vector field $\mathbf{b}(x) = (b_i(x))_{1 \le i \le k}$, $x \in \mathbb{R}^k$, such that

(i) Each component $a_{ij}(x)$ and $b_i(x)$ is a periodic function.

$$a_{ij}(x+m) = a_{ij}(x) \quad b_i(x+m) = b_i(x), \text{ for all } m \in \mathbb{Z}^k, x \in \mathbb{R}^k, 1 \le i, j \le k.$$

(ii) Each $a_{ij}(x)$ has a second order derivative.
(iii) Each $b_i(x)$ has a continuous first order derivative.

These coefficients may be associated with Kolmogorov's backward pde for the operator

$$A = \sum_{i,j} a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_i b_i(x) \frac{\partial}{\partial x_i},$$

and/or with the diffusion $X = \{X_t = (X_t^{(1)}, \ldots, X_t^{(k)}) : t \ge 0\}$ governed by the stochastic differential equation

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dB_s, \quad t > 0, \tag{17.15}$$

where $\sigma(x)$ is the positive square root of $\mathbf{a}(x)$ for each $x \in \mathbb{R}^k$. It is the latter sde perspective that is to be considered here, with the pde connection serving as a tool for some of the analysis. In particular, with some appeal to pde theory for the existence of densities, the ergodic theory associated with the diffusion X may be obtained from that of its transformation

$$\dot{X} := X \bmod 1$$

on the (compact) torus

$$\mathbb{T} = \{x \bmod 1 : x \in \mathbf{R}^k\} \equiv [0, 1)^k,$$

as follows.

Theorem 17.7 (Ergodic Theory for X Derived from $\dot{X}$) *The diffusion X has a unique invariant probability $\pi(dx)$, absolutely continuous with respect to Lebesgue measure with density $\pi(x)$. Moreover,*

$$\sup_{x \in \mathbb{T}} |p(t; x, y) - \pi(y)| \le c e^{-\beta t}, \quad t \ge 0,$$

for positive constants c, β.

Proof Recall the construction in Example 8 from Chapter 13 of a diffusion, denoted $\dot{X} = \{\dot{X}_t : t \ge 0\}$, with state space the torus $\mathbb{T}$ for the prescribed Lipschitz and periodic coefficients. Let

$$\dot{p}(t; x, dy) = \dot{p}(t; x, y)dy, \quad t > 0, x, y \in \mathbb{T},$$

denote the transition probabilities for $\dot{X}$, where

$$\dot{p}(t; x, y) = \sum_{m \in \mathbb{Z}^k} p(t; x, y + m), \quad x, y \in [0, 1)^k.$$

Since, from pde theory,[8] one has

$$\inf_{x, y \in \mathbb{T}} p(t; x, y) > 0, \quad t > 0,$$

it follows readily from (13.36) that

$$\inf_{x, y \in \mathbb{T}} \dot{p}(t; x, y) > 0, \quad t > 0,$$

as well. Moreover, $\mathbb{T}$ is compact, so Doeblin's condition holds, and the assertion of the theorem applies to $\dot{X}$, i.e., the existence of an absolutely continuous invariant probability $\pi(dx) = \pi(x)dx$ on $\mathbb{T}$ for $\dot{X}$. Now, $\pi(x)$ has a periodic extension to an invariant measure for X, again denoted by $\pi(dx) = \pi(x)dx$. ■

The main goal for this example is to prove the next theorem from Bensoussan et al. (1978) and Bhattacharya (1985). This theorem is partially motivated by an application to homogenization of dispersion in porous media according to scale, a topic to be treated in more detail in the special topics Chapter 24.

[8] See Friedman (1964), pp. 44–46.

Theorem 17.8 *Let $\{X_t : t \geq 0\}$ be a diffusion on $\mathbb{R}^k$ with differentiable periodic coefficients $b(\cdot)$ and $\sigma(\cdot)$, with period 1 (period lattice $\mathbb{Z}^k$), and generator A for the diffusion*

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t,$$

where $\{B_t : t \geq 0\}$, is a standard k-dimensional Brownian motion. That is, with $\mathbf{a}(x) = \sigma(x)\sigma(x)'$,

$$A = \sum_{1 \leq \leq j \leq k} b_j(\cdot)\frac{\partial}{\partial x_j} + \frac{1}{2}\sum_{1 \leq \leq i,j \leq k} a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j}. \tag{17.16}$$

Then

$$Z_{t,\lambda} := \lambda^{-\frac{1}{2}}(X_{\lambda t} - \lambda \overline{b} t) \Rightarrow \{Z_t : t \geq 0\}, \text{ as } \lambda \to \infty, \tag{17.17}$$

where the limiting diffusion $Z = \{Z_t : t \geq 0\}$ is a k-dimensional Brownian motion with mean zero and $k \times k$ dispersion matrix

$$K = \int_{\mathbb{T}_1} (Gradg(x) - I_k)\mathbf{a}(x)(Gradg(x) - I_k)'\pi(dx), \tag{17.18}$$

where

$$Gradg = (gradg_1, \ldots, gradg_k) \equiv (\nabla g_1, \ldots, \nabla g_k).$$

Here π is the unique invariant probability of the diffusion

$$\dot{X}_t := X_t \, mod 1, \quad t \geq 0,$$

on the unit k-dimensional torus $\mathbb{T}_1$, and $g = (g_1, \ldots, g_k)$ where g_j is the unique mean-zero periodic solution of

$$Ag_j = b_j - \overline{b}_j, \; j = 1, \ldots, k, \quad \overline{b}_j = \int_{\mathbb{T}_1} b_j(x)\pi(dx). \tag{17.19}$$

Proof By Itô's lemma, writing $\nabla = $ grad, and $\sigma_j(x)$ for the jth row of $\sigma(x)$, one has

$$\begin{aligned} & g_j(X_t) - g_j(X_0) \\ &= \int_{[0,t]} \nabla g_j(X_s)dX_s + \frac{1}{2}\sum_{1 \leq \leq i,i' \leq k} \int_{[0,t]} a_{i,i'}(X_s)\frac{\partial^2}{\partial x_i \partial x_{i'}} g_j(X_s)ds \end{aligned}$$

$$
\begin{aligned}
&= \int_{[0,t]} (b_j(X_s) - \overline{b}_j)ds + \int_{[0,t]} \nabla g_j \cdot \sigma_{j\cdot}(X_s)dB_s \\
&= X_{t,j} - X_{0,j} - t\overline{b}_j + \int_{[0,t]} \nabla g_j \cdot \sigma_{j\cdot}(X_s)dB_s - \int_{[0,t]} \sigma_{j\cdot}(X_s)dB_s, \; j = 1, \ldots, k.
\end{aligned} \tag{17.20}
$$

Hence,

$$
X_t - X_0 - t\overline{b} = g(X_t) - g(X_0) - \int_{[0,t]} (\mathrm{Grad} g(X_s) - I_k)\sigma(X_s)dB_s \tag{17.21}
$$

Under the scaling in Theorem 17.1, since

$$
\mathrm{Grad}(g(X_s)) = \mathrm{Grad}(g(\dot{X}_s))
$$

and

$$
\sigma(X_s) = \sigma(\dot{X}_s),
$$

the martingale FCLT[9] applies if $\dot{X}_0$ has the invariant distribution π. Note that g_j, ∇g_j are bounded, so that

$$
\lambda^{-\frac{1}{2}}(g(X_t) - g(X_0)) \to 0 \text{ as } \lambda \to \infty.
$$

Hence (17.16) follows from (17.20). For an arbitrary initial distribution, the FCLT still holds in view of Theorem 17.6. ■

Example 3 (Taylor-Aris Theory of Solute Transport) In a classic paper, G. I. Taylor[10] showed that when a solute in dilute concentration $c_0(dx)$ is injected into a liquid flowing with a steady flow velocity through an infinite straight capillary of uniform circular cross section, the concentration $c(t, y)$ along the capillary becomes Gaussian as time increases. Moreover the large-scale dispersion coefficient could be explicitly calculated in terms of the cross-sectional average velocity U_0, its radius $a > 0$, and the molecular diffusion coefficient D_0 as

---

[9] Bhattacharya and Waymire (2022), Theorem 15.5, or Billingsley (1968), Theorem 23.1.

[10] Taylor (1953) had argued the seminal asymptotic formula [11] $D \sim a^2U_0^2/48D_0$. The explicit computation (17.22) was given by Aris (1956) using the method of moments and eigenfunction expansions, and it is popularly called the Taylor-Aris formula.

[11] Wooding (1960) extended the Taylor-Aris analysis and formula to dispersion of a solute in a unidirectional parabolic flow between two parallel planes separated by a distance r. Accordingly, in this setting $D = 2D + \frac{8r^2U^2}{945D}$.

$$D := \frac{a^2 U_0^2}{48 D_0} + D_0. \tag{17.22}$$

In the classic treatments, the problem was viewed as that of determining the long time behavior of the solute concentration $c(t; y)$ as a solution to the Fokker-Planck equation for the concentration. Specifically, let

$$\begin{aligned} G &= \{y = (y^{(1)}, \tilde{y}) : -\infty < y^{(1)} < \infty, |\tilde{y}| < a\}, \\ \partial G &= \{y : |\tilde{y}| = a\}, \qquad \left(\tilde{y} = (y^{(2)}, y^{(3)})\right). \end{aligned} \tag{17.23}$$

The velocity of the liquid is along the capillary length (i.e., in the $y^{(1)}$ direction) and is given, as the solution of a Navier-Stokes equation governing a steady nonturbulent flow, by the parabolic flow

$$F(\tilde{y}) = U_0(1 - \frac{|\tilde{y}|^2}{a^2}). \tag{17.24}$$

The parameter U_0 is the maximum velocity, attained at the center of the capillary. Using Gauss' divergence theorem, the pde for the time evolution of the concentration from the perspective of conservation of mass (Fick's law) is of the form

$$c(t; y) = \int_G p(t; x, y) c_0(dx), \tag{17.25}$$

where, $p(t; x, y)$ satisfies the Fokker-Plank equation, i.e., Kolmogorov forward equation (see Chapter 15),

$$\frac{\partial p}{\partial t} = \frac{1}{2} D_0 \mathbf{\Delta}_y p - \frac{\partial}{\partial y^{(1)}} (F(\tilde{y}) p) \qquad (y \in G, t > 0). \tag{17.26}$$

The *forward boundary condition* is the no-flux condition

$$y^{(2)} \frac{\partial p}{\partial y^{(2)}} + y^{(3)} \frac{\partial p}{\partial y^{(3)}} \equiv \mathbf{n}(y) \cdot \mathbf{grad}\, p = 0 \qquad (y \in \partial G, t > 0). \tag{17.27}$$

It was subsequently recognized by Bhattacharya and Gupta (1984) that this problem has an equivalent formulation in terms of the injected particles. Namely, for a dilute concentration, particle motions may be assumed to be independent, and their individual evolutions simply involve an ergodic theorem for the cross-sectional dispersion in the transverse direction to the fluid flow and a central limit theorem in the longitudinal direction of the flow. Specifically, since (i) the diffusion matrix is $D_0\mathbf{I}$, (ii) the drift velocity is along the $x^{(1)}$-axis and depends only on $x^{(2)}, x^{(3)}$, and (iii) the boundary condition only involves $x^{(2)}, x^{(3)}$, the individual immersed particle trajectories $\{X_t\}$ have the following representation:

Proposition 17.9

a. $\{\tilde{X}_t := (X_t^{(2)}, X_t^{(3)})\}$ *is a two-dimensional, reflecting Brownian motion on the disc* $\overline{G}_a := \{y \in \mathbb{R}^2 : |\tilde{y}| \le a\}$ *with diffusion matrix* $D_0\mathbf{I}$, *where* $\mathbf{I}$ *is the* 2×2 *identity matrix.*

b. $X_t^{(1)} = X_0^{(1)} + \int_0^t F(\tilde{X}_s)ds + \sqrt{D_0}B_t$, *where* $B = \{B_t\}$ *is a one-dimensional standard Brownian motion, independent of* $\{\tilde{X}_t\}$ *and* $X_0^{(1)}$.

Theorem 17.10 *Let* $\tilde{X}$ *and* B *denote the independent diffusions in Proposition 17.9.*

a. $\tilde{X}$ *is an ergodic Markov process whose unique invariant probability* π *is the uniform distribution on on the disc* $\overline{G}_a$.

b. *The radial process* $\{\tilde{R}_t := |\tilde{X}_t|\}$ *is a one-dimensional diffusion on* $[0, a]$ *whose transition probability density is*

$$q(t; r, \tilde{r}) := \int_{\{\tilde{y}:|\tilde{y}|=\tilde{\mathbf{r}}\}} \tilde{p}(t; \tilde{x}, \tilde{y}) s_{\tilde{r}}(d\tilde{y}) \qquad (|\tilde{x}| = r), \tag{17.28}$$

where $s_{\tilde{r}}(d\tilde{y})$ *is the arc length measure on the circle* $\{\tilde{y} \in \mathbb{R}^2 : |\tilde{y}| = \tilde{r}\}$ *and* $p(t; \tilde{x}, \tilde{y})$ *is the transition probability of the diffusion* $\tilde{X}$ *on* $\overline{G}_a$. *Moreover,* $\{\tilde{R}_t\}$ *has the unique invariant density*

$$\pi(r) = \frac{2r}{a^2}, \qquad 0 \le r \le a, \tag{17.29}$$

and

$$\max_{0 \le r, \tilde{r} \le a} |q(t; r, \tilde{r}) - \frac{2r}{a^2}| \le 2\pi a c_1 e^{-c_2 t} \qquad (t > 0), \tag{17.30}$$

for some constants $c_1 > 0, c_2 > 0$.

c.

$$Y_t^{(n)} := n^{-1/2}(X_{nt}^{(1)} - \frac{1}{2}U_0 nt) \Rightarrow N(0, Dt), \tag{17.31}$$

where

$$D := \frac{a^2 U_0^2}{48 D_0} + D_0, \tag{17.32}$$

Proof The Markov property for these functions of X, transition probabilities, and their generators are provided in Chapter 13. The transition probability density $\tilde{p}$ of the Markov process $\{\tilde{X}_t\}$ is related to p by

$$\tilde{p}(t; \tilde{x}, \tilde{y}) = \int_{-\infty}^{\infty} p(t; x, y) dy^{(1)}. \tag{17.33}$$

The disc $\overline{G}_a$ is compact and $\tilde{p}(t; \tilde{x}, \tilde{y})$ is positive and continuous in $\tilde{x}, \tilde{y} \in \overline{G}_a$ for each $t > 0$. It follows, due to Doeblin minorization, that there is a unique invariant probability π for $\tilde{p}$ and that for some positive constants c_1, c_2, one has

$$\max_{\tilde{x}, \tilde{y} \in B_a} |\tilde{p}(t; \tilde{x}, \tilde{y}) - \pi(\tilde{y})| \leq c_1 e^{-c_2 t} \qquad (t > 0). \tag{17.34}$$

It is straightforward to check that $\pi(\tilde{y})$ is the uniform density (Exercise 6) The generator of $\{\tilde{R}_t\}$ is given by the backward operator

$$A := \frac{1}{2} D_0 (\frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr}), \; 0 < r < a, \qquad \frac{\partial}{dr}\bigg|_{r=a} = 0, \tag{17.35}$$

and its transition probability density converges exponentially fast to the invariant density (17.29). To prove (17.31), write

$$f(r) = 1 - \frac{r^2}{a^2} \qquad (0 \leq r \leq a). \tag{17.36}$$

Then

$$F(\tilde{y}) = U_0 f(|\tilde{y}|).$$

Now by the central limit theorem (Theorem 17.1), it follows that as $n \to \infty$,

$$\{Z_t^{(n)} = n^{-1/2} \int_0^{nt} (f(\tilde{R}_s) - \mathbb{E}_\pi f) ds : t \geq 0\} \Rightarrow \sigma B, \tag{17.37}$$

where

$$\mathbb{E}_\pi f = \int_0^a f(r)\pi(r) dr = \int_0^a (1 - \frac{r^2}{a^2})(\frac{2r}{a^2}) dr = \frac{1}{2}. \tag{17.38}$$

The key calculation of the variance parameter σ^2 of the limiting Gaussian yields

$$\sigma^2 = 2 \int_0^a (f(r) - \mathbb{E}_\pi f) h(r) \pi(r) dr \tag{17.39}$$

where h is a function in $L^2([0, a], \pi)$ satisfying

$$Ah(r) = -(f(r) - \mathbb{E}_\pi f) = \frac{r^2}{a^2} - \frac{1}{2}. \tag{17.40}$$

Such an h is unique up to the addition of a constant and is easily given by

$$h(r) = \frac{1}{8D_0}(\frac{r^4}{a^2} - 2r^2) + c_3 \tag{17.41}$$

where c_3 is an arbitrary constant, which may be taken to be zero in carrying out the integration in (17.39). One then obtains

$$\sigma^2 = \frac{a^2}{48D_0}. \tag{17.42}$$

Finally, since $\{B_t\}$ and $\{\tilde{X}_t\}$ in part (b) of Proposition 17.9 are independent, the formula for D follows. ■

Remark 17.8 (17.31) and (17.32) represent Taylor's main result as completed by Aris (1956). It is noteworthy that the effective dispersion D is larger than the molecular diffusion coefficient D_0 and grows quadratically with velocity.

Remark 17.9 That the probabilisitic method is truly different from the use of eigenfunctions by Aris (1956) is shown in Bhattacharya and Gupta (1984) by deriving an identity involving the zeros of Bessel functions of order one by comparing the two methods and even conjecturing a result which seems to be new.

Remark 17.10 The problem of extending these results to cover nonsmooth dispersion coefficients at the molecular scale arose in hydrologic experiments involving sharp heterogeneities in the subsurface medium. The conventional wisdom was that the large-scale dispersion would still be Gaussian with a dispersion coefficient of the Taylor-Aris general form (17.22), but involving 'averaged'values of the molecular scale diffusion coefficient. In an analysis of this problem, Ramirez et al. (2006) showed that the probabilistic approach[12] of Bhattacharya and Gupta (1984), together with properties of skew Brownian motion, leads to Gaussian dispersion, with an explicit formula for the large-scale dispersion. In particular the specific nature of the asymptotic dispersion is shown to involve both *arithmetic and harmonic averaging* of the molecular scale diffusion coefficient in the formula for the large-scale dispersion.

Remark 17.11 (A Gaussian Random Field Indexed By $\mathcal{R}_A$.) It follows immediately from Theorem 17.1 that one has the mean-zero Gaussian random field $\mathscr{G}$ indexed by $\mathcal{R}_A$, where

$$\mathscr{G} := \{Z(f) : f \in \mathcal{R}_A\},$$

[12] A technical issue occurs in that the independence of $\{B_t\}$ and $\{\tilde{X}_t\}$ in Proposition 17.9(b) no longer holds for finite time t.

with covariance

$$\mathrm{Cov}(Z(f_1), Z(f_2)) = -\langle f_1, g_2 \rangle - \langle f_2, g_1 \rangle,$$

where g_i, $i = 1, 2$, is the unique mean-zero element of $\mathcal{D}_A$ such that $Ag_i = f_i$, $i = 1, 2$.

Exercises

1. (a) Show that the adjoint operator A^* is well defined if D_A is dense in $L^2(S, \pi)$. What is the domain of A^*? [*Hint*: Let $g \in L^2$ be such that the map $f \to \langle Af, g \rangle$ is continuous on $\mathcal{D}_A$. Then it has a unique extension to L^2, and there exists a unique element A^*g, say, such that $\langle A^*g, f \rangle = \langle Ag, f \rangle$. $\mathcal{D}_{A^*}$ comprises all such g and A^*g is the corresponding linear functional as defined.] (b) Prove that $\mathcal{R}_A \subset \mathcal{N}_A^{\perp}$. [*Hint*: Use Theorem 17.1(c) and Proposition 17.2.]
2. In the proof of Proposition 17.4, show that $g(X_t) - g(X_0) = 0$ π-a.s.. [*Hint*: Under the hypothesis, the martingale differences $\Delta_n g$ are zero π-a.s.. Hence one has the identity $g(X_h) - \int_0^n f(X_s)ds = g(X_0)$. The martingale on the left side does not depend on n. It is uniformly integrable, and its limit, as $n \to \infty$, is the same as its value for every integer $n > 0$. The same arguments show that $g(X_h) - \int_m^n f(X_s)ds = g(X_m)$ for all $m > n$. But the limit of the martingale on the left side is the same as that of the preceding identity. Hence the two right sides must coincide, i.e., $g(X_m) = g(X_0)$ a.s..]
3. Supply the details to prove (17.13).
4. Consider a positive recurrent diffusion on $\mathbb{R}$ (see Chapters 11 and 21). (a) Show that $\mathcal{R}_A = 1^{\perp}$. (b) Show that $f \in \mathcal{R}_A$ if and only if $x \to \int_0^x \int_{-\infty}^y f(z)m(dz)dy \in L^2(\mathbb{R}, \pi)$, where π is the unique invariant probability. (c) Write out the density of π with respect to Lebesgue measure. [*Hint*: See Chapter 11.] (d) Check that for $f \in \mathcal{R}_A$, $\sigma^2 = -2\langle f, g \rangle = \frac{2}{m(\mathbb{R})} \int_{-\infty}^{\infty} f(x)\{\int_x^0 dy \int_{-\infty}^y f(z)dz\}m(dx)$. (e) Show that one needs to assume $f \in 1^{\perp}$, and $x \to \int_0^x \{\int_{(-\infty,y]} f(z)m(dz)\}dy \in L^2$.
5. Write out the asymptotic variance for Theorem 17.8 with $k = 1$, i.e., $\mathbb{T}_1 = \{x \bmod 1 : x \in \mathbb{R}\}$, which may be viewed as the unit circle in the complex plain, $\{e^{2\pi i x} : 0 \le x \le 1\}$.
6. Show that the invariant distribution for (17.33) in the Taylor-Aris example is uniform. [*Hint*: Use the backward equation (and boundary condition) for $\tilde{p}$ to show that $\frac{\partial}{\partial t} \int_{\{|\tilde{y}| \le a\}} \tilde{p}(t; \tilde{x}, \tilde{y})\pi(\tilde{x})d\tilde{x} = 0$.]
7. Wooding (1960) Compute the asymptotic formula $D = 2D + \frac{8r^2U^2}{945D}$ for dispersion of a solute in a unidirectional parabolic flow between two parallel planes separated by a distance r.

Chapter 18
Asymptotic Stability for Singular Diffusions

In the present chapter, asymptotic stability conditions for diffusions with linear drift and Lipschitz diffusion coefficients are presented. Although originally designed to provide criteria for the stability in distribution of multidimensional diffusions with linear drift and singular diffusion coefficients, the method presented may be used for nonsingular diffusions as well and also for some nonlinear drift.

In this chapter we consider diffusions $\{X^x(t) : t \geq 0\}$ on $\mathbb{R}^k$ defined as solutions to the stochastic differential equation

$$X^x(t) = x + \int_0^t BX^x(s)ds + \int_0^t \sigma(X^x(s))dW(s), \quad t \geq 0, \tag{18.1}$$

where B is a $k \times k$ matrix, $\sigma(\cdot)$ is a Lipschitz $k \times \ell$-matrix valued function on $\mathbb{R}^k$, and $W = \{W(t) : t \geq 0\}$ is a ℓ-dimensional standard Brownian motion. In particular, assume there is a $\lambda_0 \geq 0$ such that

$$||\sigma(x) - \sigma(y)|| \leq \lambda_0 |x - y|, \quad x, y \in \mathbb{R}^k, \tag{18.2}$$

where $||\cdot||$ denotes matrix norm with respect to the Euclidean norm $|\cdot|$. The transition probability of the diffusion will be denoted $p(t; x, dy)$. The goal is to provide conditions on B and $\sigma(\cdot)$ to ensure *stability in distribution* in the sense that there is a probability measure $\pi(dy)$ such that, for every initial state x, $p(t; x, dy)$ converges weakly to $\pi(dy)$ as $t \to \infty$.

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_18

Criteria were established in Chapter 11 in the case that the matrices $\sigma(\cdot)\sigma(\cdot)^t$ are nonsingular. In such (nonsingular) cases, asymptotic stability is equivalent to the existence of a unique invariant probability. The emphasis here is on singular diffusions, although the method applies to nonsingular diffusions also. There are two components of the present method: (1) tightness and (2) asymptotic flatness defined below.

Example 1 Let $k = 2$ and take

$$B = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \sigma(x) = c \begin{pmatrix} x_2 & 0 \\ -x_1 & 0 \end{pmatrix}.$$

Then $R^2(t) = X_1^2(t) + X_2^2(t), t \geq 0$, satisfies

$$dR^2(t) = (c^2 - 2)R^2(t)dt,$$

so that

$$R^2(t) = R^2(0)e^{(c^2-2)t}, \quad t \geq 0.$$

In the case $|c| = \sqrt{2}$, $R^2(t) = R^2(0)$ for all $t \geq 0$. Thus, the transition probabilities $\{p(t; x, dy) : t \geq 0\}$ of $\{X^x(t) : t \geq 0\}$ are tight for every x. However, there is an invariant probability on every circle, and the angular motion on the circle is a periodic diffusion. On the other hand, if $c \neq \sqrt{2}$, then the only invariant probability is the point mass at the origin, but if $|c| > \sqrt{2}$, and $X(0) \neq 0$, then $R^2(t) \to \infty$ a.s. as $t \to \infty$. If $|c| < \sqrt{2}$, then the diffusion is stochastically stable a.s. and, therefore, in distribution.

In the present chapter, stability criteria[1] that will apply to both nonsingular and singular diffusions are presented, i.e., for the latter, cases in which $\sigma(x)\sigma(x)^t$ is of rank less than the dimension k for some x.

Lemma 1 below provides an approach to this problem. It involves viewing the space-time family $\{X^x(t) : t \geq 0, x \in \mathbb{R}^k\}$ as a *stochastic flow* in the spirit of dynamical systems. That is, one may informally view it as a collection of time evolutions of particles originating at the points $x \in \mathbb{R}^k$.

Definition 18.1 A stochastic flow $\{X^x(t) : t \geq 0, x \in \mathbb{R}^k\}$ is said to be *asymptotically flat uniformly on compacts* (in probability) if for every $\varepsilon > 0$, compact $K \subset \mathbb{R}^k$, one has

$$\sup_{x,y \in K} P(|X^x(t) - X^y(t)| > \varepsilon) \to 0 \text{ as } t \to \infty. \tag{18.3}$$

[1] The results presented here were originally obtained by Basak and Bhattacharya (1992).

Lemma 1 *The following two conditions are sufficient to imply stochastic stability: (i) $\{p(t; x, dy) : t \geq 0\}$ is tight, and (ii) the stochastic flow is asymptotically flat uniformly on compacts.*

Proof For each x tightness provides a probability measure $\pi^x(dy)$, say, such that $p(t_{m(x)}; x, dy) \Rightarrow \pi^x(dy)$ as $m \to \infty$ for an unbounded sequence $0 \leq t_{1(x)} < t_{2(x)} < \cdots$. For $z \in \mathbb{R}^k$, tightness of $\{p(t_{m(x)}; z, dy) : m \geq 1\}$ provides a common subsequence, say $\{t_m\}$, such that $p(t_m; x, dy) \Rightarrow \pi^x$, and $p(t_m; z, dy) \Rightarrow \pi^z$ as $m \to \infty$. Suppose $\pi^x \neq \pi^z$. Then there is a continuous function g with compact support, say K, such that

$$\int_{R^k} g(y)\pi^x(dy) \neq \int_{\mathbb{R}^k} g(y)\pi^z(dy), \tag{18.4}$$

but

$$\lim_{m\to\infty} \int_{\mathbb{R}^k} g(y)p(t_m; x, dy) = \lim_{m\to\infty} \int_{\mathbb{R}^k} g(y)p(t_m; z, dy).$$

So (18.4) contradicts the assumption of asymptotic flatness uniformly on compacts. ■

In order to establish asymptotic flatness uniformly on compacts, a stronger property of *asymptotic flatness in the δ-th mean*, which means for $\delta > 0$, and compact $K \subset \mathbb{R}^k$,

$$\lim_{t\to\infty} \sup_{x,y\in K} \mathbb{E}|X^x(t) - X^y(t)|^\delta = 0. \tag{18.5}$$

The following example shows that asymptotic flatness uniformly on compacts, alone, is far from being sufficient for stability.

Example 2 Let $k = 1$, $b(x) = e^{-x}$, $\sigma(x) \equiv 0$. Then $X^x(t) = \log(t + e^x) \to \infty$ as $t \to \infty$, but

$$X^x(t) - X^z(t) = \log \frac{t + e^x}{t + e^z} \to 0$$

as $t \to \infty$, uniformly for x, z in a compact set K.

Let

$$a(x) = \sigma(x)\sigma(x)', \; a(x, y) = (\sigma(x) - \sigma(y))(\sigma(x) - \sigma(y))'. \tag{18.6}$$

The main result for this chapter can be stated as follows.

Theorem 18.1 *Assume the Lipschitz condition (18.2).*

a. *Assume there is a symmetric positive definite matrix C and a positive constant γ such that for $x \neq y$,*

$$2C(x-y)\cdot B(x-y) - \frac{2C(x-y)\cdot a(x,y)C(x-y)}{(x-y)\cdot C(x-y)} + tr(a(x,y)C) \leq -\gamma|x-y|^2, \tag{18.7}$$

where $\cdot$ is the Euclidean dot product and tr denotes the matrix trace. Then the diffusion $\{X^x(t) : t \geq 0\}$ is stable in distribution.

b. *Assume there is a symmetric positive definite matrix C and constant $\beta > 0$ such that for all sufficiently large $|x|$,*

$$2Cx\cdot Bx - \frac{2Cx\cdot a(x)Cx}{x\cdot Cx} + tr(a(x)C) \leq -\beta|x|^2. \tag{18.8}$$

Then an invariant probability exists.

Corollary 18.2 *Assume the Lipschitz condition (18.2). Then (18.7) is sufficient for stability.*

Proof This follows since for Lipschitz $\sigma(\cdot)$, (18.2) implies (18.8) for every $\beta \in (0, \lambda)$, as can be seen as follows. Take $y = 0$ in (18.7) and note that

$$\begin{aligned} -\gamma|x|^2 &\geq 2Cx\cdot Bx - \frac{2Cx\cdot a(x,0)Cx}{x\cdot Cx} + \mathrm{tr}(a(x,0)C) \\ &= 2Cx\cdot Bx - \frac{2Cx\cdot a(x)Cx}{x\cdot Cx} + \mathrm{tr}(a(x)C) + O(|x|), \text{ as } |x| \to \infty. \end{aligned}$$

■

In the course of the proof of Theorem 18.1 below, one sees that (18.8) implies the *existence* but not the *uniqueness* implied by stability. The stronger condition (18.7) also implies asymptotic flatness (18.3). The existence of an invariant probability and asymptotic flatness together immediately yield uniqueness and stability, as demonstrated by the proof of Lemma 1.

The following Corollary 18.3 may be viewed as a generalization of a well-known necessary and sufficient condition[2] on the matrix B for stability in the case of a constant matrix $\sigma(\cdot) \equiv \sigma$. Also see Exercise 6.

Corollary 18.3 *Assume that $\sigma(\cdot)$ is Lipschitz with coefficient λ_0 and all eigenvalues of B have negative real parts. Assume in addition that*

$$(k-1)\lambda_0^2 < \frac{1}{\Lambda_P}, \tag{18.9}$$

where Λ_P is the largest eigenvalue of

[2] See Arnold (1974), pages 178–187.

$$P = \int_0^\infty \exp(sB') \exp(sB) ds. \tag{18.10}$$

The diffusion $\{X^x(t) : t \geq 0\}$ *is stable in distribution.*

Proof Let $C = P$. Then,

$$B'P + PB = -\mathbf{I}_k, \tag{18.11}$$

where $\mathbf{I}_k$ is the $k \times k$ identity matrix (Exercise 2). Thus, writing "tr" for trace,

$$2Px \cdot Bx = x \cdot (PB + B'P)x = -|x|^2,$$

$$\text{tr}(a(x, y)P) = tr(\sqrt{P}a(x, y)\sqrt{P}),$$

and

$$\frac{2P(x-y) \cdot a(x, y)P(x-y)}{(x-y) \cdot P(x-y)}$$
$$= \frac{2\sqrt{P}(x-y) \cdot \sqrt{P}a(x, y)\sqrt{P}\sqrt{P}(x-y)}{\sqrt{P}(x-y) \cdot \sqrt{P}(x-y)}$$

$$\geq 2(\text{smallest eigenvalue of } \sqrt{P}a(x, y)\sqrt{P}).$$

Thus, noting that $\text{tr}a(x, y)P = \text{tr}a(x, y)\sqrt{P}\sqrt{P} = \text{tr}\sqrt{P}a(x, y)\sqrt{P}$, one has

$$\frac{2P(x-y) \cdot a(x, y)P(x-y)}{(x-y) \cdot P(x-y)} + \text{tr}(a(x, y)P)$$
$$\leq \sum_{j=2}^{k} \lambda_j(x, y) - \lambda_1(x, y), \tag{18.12}$$

where $\lambda_1(x, y) \leq \lambda_2(x, y) \leq \cdots \leq \lambda_k(x, y)$ are the eigenvalues of $\sqrt{P}a(x, y)\sqrt{P}$. Now, the right side of (18.12) is no larger than

$$(k-1)||\sqrt{P}a(x, y)\sqrt{P}|| \leq (k-1)||P||(||a(x, y)||) \leq (k-1)\Lambda_P \lambda_0^2 |x-y|^2.$$

Now take $\gamma = (1 - (k-1)\Lambda_P \lambda_0^2)$ to obtain (18.7). ■

Proof of Theorem 18.1 Consider the Lyapunov[3] function

[3] The term Lyapunov function is used here in the general sense of their use for proving stability of dynamical systems.

$$v(x) = (x \cdot Cx)^{1-\varepsilon}, \tag{18.13}$$

with $\varepsilon \in [0, 1)$ to be determined. Define, for a given pair (x, y), $x \neq y$, define

$$\begin{aligned} Z^{x,y}(t) &:= X^x(t) - X^y(t) \\ &= x - y + \int_0^t BZ^{x,y}(s)ds + \int_0^t (\sigma(X^x(s)) - \sigma(X^y(s)))dW(s), \end{aligned} \tag{18.14}$$

and let

$$\tau_0 = \inf\{t \geq 0 : Z^{x,y}(t) = 0\}.$$

By Itô's lemma, denoting $\partial_i = \frac{\partial}{\partial x_i}$, one has for $t < \tau_0$ (see Theorem 8.4),

$$\begin{aligned} &v(Z^{x,y}(t)) - v(x-y) \\ &= \int_0^t \tilde{L}(v)(X^x(s), X^y(s))ds + \int_0^t (\mathrm{grad} v)(Z^{x,y}(s)) \cdot (\sigma(X^x(s)) - \sigma(X^y(s))dW(s), \end{aligned} \tag{18.15}$$

where, using (18.15), $\mathrm{grad}(x-y) \cdot C(x-y) = 2C(x-y)$, $\partial_i \partial_j (x-y) \cdot C(x-y) = 2c_{ji} = 2c_{ij}$, and $\sum_{i,j} a_{ij} c_{ji} = \mathrm{tr} a(x, y)C$,

$$\begin{aligned} &\tilde{L}(v)(x, y) \\ &\quad := B(x-y) \cdot (\mathrm{grad} v)(x-y) + \frac{1}{2} \sum_{i,j=1}^{k} a_{ij}(x, y)(\partial_i \partial_j v)(x-y) \\ &\quad = (1-\varepsilon)((x-y) \cdot C(x-y))^{-\varepsilon} \{2B(x-y) \cdot C(x-y) \\ &\qquad -2\varepsilon \frac{(x-y) \cdot Ca(x, y)C(x-y)}{(x-y) \cdot C(x-y)} + \mathrm{tr}(a(x, y)C)\} \\ &\quad \leq (1-\varepsilon)((x-y) \cdot C(x-y))^{-\varepsilon} \{-\gamma |x-y|^2 \\ &\qquad +2(1-\varepsilon) \frac{(x-y) \cdot Ca(x, y)C(x-y)}{(x-y) \cdot C(x-y)} \\ &\leq (1-\varepsilon)((x-y) \cdot C(x-y))^{-\varepsilon} \times \{-\gamma |x-y|^2 + 2(1-\varepsilon)\lambda_0^2 \Lambda_C |x-y|^2\}. \end{aligned} \tag{18.16}$$

Here Λ_C is the largest eigenvalue of C. Now choose $\varepsilon \in [0, 1)$ such that

$$-\gamma_1 := -\gamma + 2(1-\varepsilon)\lambda_0^2 \Lambda_C < 0. \tag{18.17}$$

Then we have

$$\tilde{L}(v)(x, y) \leq -\alpha v(x - y), \tag{18.18}$$

where $\alpha := \frac{\gamma_1(1-\varepsilon)}{\Lambda_C}$. Consider the process

$$Y(t) := e^{\alpha t} v(Z^{x,y}(t)), \quad t \geq 0.$$

It follows from (18.15) and (18.18) that $\{Y(t \wedge \tau_0) : t \geq 0\}$ is a positive supermartingale. Thus,

$$\mathbb{E}Y(t \wedge \tau_0) \leq \mathbb{E}Y(0) = v(x - y). \tag{18.19}$$

Since $Z^{x,y}(t) = 0$ on $[\tau_0 \leq t]$, $Y(t) = 0$ on $[\tau_0 \leq t]$. (18.19) implies $\mathbb{E}Y(t) \leq v(x - y)$. That is,

$$\mathbb{E}(Z^{x,y}(t) \cdot CZ^{x,y}(t))^{1-\varepsilon} \leq e^{-\alpha t}((x - y) \cdot C(x - y))^{1-\varepsilon}, \quad t \geq 0. \tag{18.20}$$

This establishes the asymptotic flatness of the stochastic flow (in the $2(1 - \varepsilon)$-th mean). To complete the proof, we need to check tightness. Let L denote the infinitesimal generator of $\{X_t^x : t \geq\}$. Modify v near the origin to be twice continuously differentiable on $\mathbb{R}^k$ if necessary. Then (18.8) implies that

$$Lv(x) \to -\infty \text{ as } |x| \to \infty.$$

We use Itô's lemma to check that this implies the existence of an invariant probability to complete the proof. Let

$$M = \sup_{x \in \mathbb{R}^k} Lv(x),$$

$$M_n = - \sup_{|x| > N} Lv(x), n = 1, 2, \ldots.$$

Then, $M_n \to \infty$ as $n \to \infty$, and

$$\begin{aligned}
\mathbb{E}v(X_t^x) &= v(x) + \mathbb{E}\int_0^t Lv(X_s^x))ds \\
&= v(x) + \mathbb{E}\int_0^t \mathbf{1}_{[|X_s^x| > n]} Lv(X_s^x))ds \\
&\quad + \mathbb{E}\int_0^t \mathbf{1}_{[|X_s^x| \leq n]} Lv(X_s^x))ds \\
&\leq -M_n \mathbb{E}\int_0^t \mathbf{1}_{[|X_s^x| > n]} ds + Mt
\end{aligned}$$

$$= -M_n \int_0^t p(s; x, \overline{B}_n^c) ds + Mt + v(x), \tag{18.21}$$

where the last inequality is by Fubini-Tonelli, and $B_n = \{y : |y| < n\}$. Thus,

$$\frac{1}{t} \int_0^t p(s; x, \overline{B}_n^c) ds \leq \frac{M}{M_n} + \frac{v(x) - \mathbb{E}v(X_t^x)}{M_n t},$$

and therefore

$$\frac{1}{t} \int_0^t p(s; x, \overline{B}_n^c) \to 0 \text{ as } n \to \infty.$$

Thus, given $\delta > 0$, there is a compact set $K_\delta = \overline{B}_n$ for sufficiently large n, such that

$$\frac{1}{t} \int_0^t p(s; x, K_\delta) ds > 1 - \delta, \text{ for all } t > 0.$$

Thus, choose an unbounded, increasing sequence of times $t_n, n \geq 1$, such that $\{p(t_n; x, dy)\}$ is tight. Now apply Lemma 1 to prove part (b), and complete the proof of Theorem 18.1. ■

For nonsingular diffusions, the Khaśminskii criterion for positive recurrence given in Chapter 11 may be applied, but it is not very suitable if the infinitesimal generator is far from being radial. In the following example, Khaśminskii criterion is not satisfied but (18.8) holds.

Example 3 Let

$$k = 2, \quad B = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad a(x) = (\delta_1 + \delta_2 x_2^2) I_2,$$

for constants $\delta_1 > 0$, $\delta_2 \geq 0$. Recall, for nonsingular $\sigma(\cdot)$, the Khaśminskii criterion for stability, given by Theorem 11.8, computes as follows:

$$\int_1^\infty e^{-\overline{I}(r)} dr = \infty \text{ for all } \delta_2 \geq 0$$

$$\int_1^\infty \frac{e^{\overline{I}(r)}}{\underline{\alpha}(r)} dr < \infty \text{ if and only if } \delta_2 < 1. \tag{18.22}$$

In particular, according to this test, the diffusion is stable in distribution if $\delta_2 \in [0, 1)$. On the other hand, taking $C = I_2$, the left side of (18.8) is $-2|x|^2$. Thus the criterion (18.8) is satisfied, and the diffusion is stable in distribution regardless of the value of $\delta_2 \geq 0$.

It is natural to ask[4] if a diffusion on $\mathbb{R}^k$ would be stable in distribution if $\sigma(\cdot)$ is Lipschitz, $\sigma(0) \neq 0$, and all eigenvalues of B have negative real parts. Equation (18.9) shows that the answer is affirmative for $k = 1$. Example 1 provides a counterexample in the case $k = 2$, with $\sigma(\cdot)$ modified near the origin. The following two examples show that even when restricted to nonsingular diffusions, the answer is "yes"for $k = 1$, and "no"for $k > 1$.

Example 4 Let

$$k > 2, \quad B = -\delta \mathbf{I}_k \ (\delta > 0), \quad a(x) = dr^2 \mathbf{I}_k, \ (d > 0, r = |x| \geq 1). \tag{18.23}$$

Then $a(\cdot)$ is nonsingular and Lipschitz on $\mathbb{R}^k$. In this case, Khaśmininskii's criterion is *necessary* as well as sufficient, and

$$\overline{I}(r) = (k - 1 - 2\frac{\delta}{d}) \ln r, \quad r \geq 1.$$

If $\delta/d < (k-2)/2$, then the first integral in (18.22) converges, implying that the diffusion is transient. If $\delta/d = (k-2)/2$, then the first integral in (18.22) diverges, as does the second integral. Thus the diffusion is null-recurrent. If $\delta/d > (k-2)/2$, then (18.22) holds and the diffusion is, therefore, positive recurrent and hence stable in distribution.

Example 5

$$k = 2, \quad B = -\delta I_2 \ (\delta > 0), \quad a(x) = \begin{pmatrix} \lambda_1 x_2^2 + \varepsilon & -\lambda_1 x_1 x_2 \\ -\lambda_1 x_1 x_2 & \lambda_1 x_1^2 + \varepsilon \end{pmatrix},$$

where δ, λ_1, and ε are positive constants. Note that the positive definite square root of $a(x)$ is Lipschitz in this case. Khaśminskii criterion for recurrence (or transience) is necessary and sufficient here, and computes as follows:

$$\begin{aligned} \overline{\beta}(r) = \underline{\beta}(r) &:= \inf_{|x|=r} \frac{(\lambda_1 - 2\delta)|x|^2 + \varepsilon}{\varepsilon} = 1 + \frac{\lambda_1 - 2\delta}{\varepsilon} r^2, \\ \overline{I}(r) = \underline{I}(r) &:= \int_1^r \frac{\underline{\beta}(u)}{u} du. \end{aligned} \tag{18.24}$$

If $\lambda_1 = 2\delta$, then both integrals in (18.22) diverge, which implies that the diffusion is null recurrent. If $\lambda_1 > 2\delta$, then the first integral in (18.2) is finite, so that the diffusion is transient. In particular, there is no invariant probability in this case.

[4] This question was posed by L. Stettner during a 1986-IMA workshop at the University of Minnesota.

Example 6 In Theorem 18.1, if $\sigma(\cdot)$ is a constant matrix, singular or nonsingular, then $a(x, y)$ vanishes. Hence if all the eigenvalues of B have negative real parts, then the diffusion has a unique invariant probability and it is stable in distribution. (See Exercise 5).

Remark 18.1 It should be noted that Theorem 18.1 holds with Bx replaced by $b(x)$, if the inequality (18.7) holds. There is no change in the proof. The problem of course is to show that the inequality (18.7) holds for a given $b(\cdot)$.

Example 7 (Nonlinear Drift) Let $b(x) = (b_1(x), \dots, b_k(x))$. Suppose $b_j(x)$ is a polynomial in x_j alone, involving only odd powers, apart from a constant term:

$$b_j(x_j) = c_{j,0} + c_{j,1}x_j + \cdots + c_{j,m(j)}x_j^{m(j)},$$

where $m(j) \geq 1$, $c_{j,1} < 0$, and $c_{j,3}, \dots, c_{j,m(j)} \leq 0$, for all $j = 1, \dots, k$. If $\sigma(\cdot)$ is a constant matrix, then (18.7) holds with $C = \mathbf{I}_k$, and the diffusion is stable in distribution. If $\sigma(\cdot)$ is Lipschitz: $|\sigma(x) - \sigma(y)| \leq \lambda_0|x - y|$, then (18.7) holds if $2\lambda > (k-1)\lambda_0^2$, where $\lambda = \min\{|c_{j,1}| : j = 1, \dots, k\}$ (Exercise 7).

Example 8 (Nonlinear Drift) Consider a nonlinear drift $b(\cdot)$. Assuming $b(\cdot)$ is continuously differentiable, one has the identity

$$b(x) - b(y) = \int_{[0,1]} \frac{d}{d\theta}(b(y + \theta(x-y))d\theta = \int_{[0,1]} J(y + \theta(x-y))(y)(x-y)d\theta,$$

where $J(x) = \text{grad}b(x) \equiv ((\partial b_i(x)/\partial x_j))$. If C is a symmetric positive definite matrix, then

$$\begin{aligned}
&C(x-y) \cdot (b(x) - b(y)) \\
&\quad = \int_{[0,1]} C(x-y)J(y + \theta(x-y))(x-y)d\theta \\
&\quad = \int_{[0,1]} J'(y + \theta(x-y))C(x-y) \cdot (x-y)d\theta \\
&\quad = \frac{1}{2}\int_{[0,1]} ((CJ + J'C)(y + \theta(x-y)(x-y) \cdot (x-y))d\theta. \qquad (18.25)
\end{aligned}$$

Assume now that $\frac{1}{2}(CJ + J'C)(\cdot)$ has all its eigenvalues less than or equal to $-\lambda < 0$. Then, if $\sigma(\cdot)$ is constant, (18.7) holds, and the diffusion with drift $b(\cdot)$ has a unique invariant measure, and it is stable in distribution. If, instead, $\sigma(\cdot)$ is Lipschitz: $|\sigma(x) - \sigma(y)| \leq \lambda_0|x - y|$, then, as in the proof of Theorem 18.1, the diffusion is stable in distribution if $2\lambda > (k-1)\lambda_0^2||C||$. In the special case $b(x) = Bx$ considered in Corollary 18.3, namely, B has all its eigenvalues with negative real parts, one may apply the above with $C = P$ (see (18.10)), to get $(CJ + J'C) = (PB + B'P) = -\mathbf{I}_k$, i.e., $2\lambda = 1$ (see (18.11)). As pointed out in Example 6, in this

case, if $\sigma(\cdot)$ is a constant matrix, the diffusion is stable in distribution. In the case $\sigma(\cdot)$ is Lipschitz, it follows from the above that the diffusion is stable in distribution if $1 > (k-1)\lambda_0^2||P||$, the same result as Corollary 18.3. Example 6 is a special case of this method, with J diagonal. It is usually relatively simple to deal with dimension $k = 2$, since one can compute the eigenvalues by solving a quadratic equation.(See Exercise 8).

Remark 18.2 In the case $\sigma(\cdot)$ is nonsingular and Lipschitz, one only needs to check the simpler condition (18.8), in order to deduce stability in distribution. This follows from Chapter 11, i.e., one only needs to prove tightness of the family $\{p(t; x, dy) : t > 0\}$ of probability measures, for some x.

We close with the following interesting example introduced by Kolmogorov to further illustrate the scope of possibilities.

Example 9 (Kolmogorov's Nonsingular Diffusion with Singular Diffusion Coefficient) This example of Kolmogorov shows that a d-dimensional diffusion with a singular diffusion term of rank less than d can still be nonsingular, i.e., have a d-dimensional smooth density everywhere. Consider the two-dimensional diffusion defined by

$$dV_t = \beta V_t dt + \sigma dB_t, \quad dX_t = V_t dt,$$

where σ is a nonzero constant and $\{B_t : t \geq 0\}$ is a one-dimensional standard Brownian motion. Then $\{V_t : t \geq 0\}$, starting at any arbitrary state, is a one-dimensional Ornstein-Uhlenbeck process, in particular with a unique invariant Gaussian density if $\beta < 0$. The position process $\{X_t : t \geq 0\}$ is also Gaussian, and the joint distribution of $\{(X_t, V_t) : t \geq 0\}$ has a positive two-dimensional Gaussian density for $t > 0$, regardless of the initial state $(V_0, X_0) = (v_0, x_0)$. Hormander's profound *Hypoellipticity Theorem* says that if the Lie algebra defined by certain $d + 1$ smooth vector fields based on the drifts and diffusion coefficients is of rank d at every point of $\mathbb{R}^d$, then the d-dimensional diffusion has a smooth positive joint density everywhere.[5]

Exercises

1. Consider Example 1. (a) Let $|c| = \sqrt{2}$. Show that there is an invariant probability on every circle $\{x : |x| = r\}$. (b) Let $c \neq \sqrt{2}$. Show that the only invariant probability is $\delta_{\{0\}}$. Show also that there is no stabilizing distribution if $|c| > \sqrt{2}$ but that there is a stabilizing distribution if $|c| < \sqrt{2}$. (c) Let $|c| = \sqrt{2}$, but

[5] For a precise statement and derivation of an extension of Hormander's result using Malliavin calculus, we refer to Chapter 5 of Ikeda and Watanabe (1989).

modify $\sigma(\cdot)$ to be nonsingular on $\{x : |x| < r_0\}$ for some $r_0 > 0$, and keep it the same as in the example for $|x| > r_0$. Investigate the asymptotics under this change.

2. (a) Prove $e^{A+B} = e^A e^B$ are finite (convergent) and A, B commute. Prove (18.11). [*Hint*: $B'e^{s(B'+B)} + e^{s(B'+B)}B = \frac{d}{ds}e^{s(B'+B)}$, so that $\int_0^\infty \{B'e^{s(B'+B)} + e^{s(B'+B)}B\}ds = e^{s(B'+B)}|_0^\infty = -\mathbf{I}_k$.]
3. Show that stability in distribution implies that π is the unique invariant probability.
4. (*Stochastic Stability of a Trap*) If the drift is a Lipschitz function $b(\cdot)$, i.e., not necessarily linear, then a point x^* such that $b(x^*) = \sigma(x^*) = 0$, is referred to as a *trap*. Show that if x^* is a trap then a sufficient condition that $X_t^x \to x^*$ a.s. and in the δ-th mean, exponentially fast for every x as $t \to \infty$, is given by $2C(x-x^*)\cdot b(x) - \frac{2C(x-x^*)\cdot a(x)C(x-x^*)}{(x-x^*)\cdot C(x-x^*)} + \mathrm{tr}(a(x,x^*)C) \le -\gamma|x-x^*|^2$ for all $x \ne x^*$.
5. Assume that $\sigma(\cdot) = \sigma$ is a constant matrix.
   (i) Use the Picard method of successive approximation to show that $X_t^x = e^{tB}x + \int_0^t e^{(t-s)B}\sigma dW(s), t \ge 0$.
   (ii) Use Itô's Lemma to verify this solution.
   (iii) Show that
   $$\Sigma(t) = \mathbb{E}\Big(\int_0^t e^{(t-s)B}\sigma dW(s)\Big)\Big(\int_0^t e^{(t-s)B}\sigma dW(s)\Big)' = \int_0^t e^{uB}\sigma\sigma' e^{uB}du.$$
   (iv) Assume that the real parts of the eigenvalues $\lambda_j, j = 1, \ldots, k$ are negative. Show that $\Sigma(t) \to \Sigma := \int_0^\infty e^{uB}\sigma\sigma' e^{uB}du$ as $t \to \infty$. [*Hint*: The modulus of an eigenvalue of e^B is $|e^{\lambda_j}| < 1, 1 \le j \le k$, and therefore $e^{tB}x \to 0$ as $t \to \infty$.]
   (v) Show that $\{X_t^x : t \ge 0\}$ is stable with Gaussian invariant probability having mean zero and dispersion matrix Σ.
   (vi) Take $B = -\delta\mathbf{I}_k$ where $\delta > 0$ Compute the dispersion matrices $\Sigma(t)$, Σ explicitly in this case, and show that they are singular if and only if σ is singular.
6. Give an example of a 2×2 matrix B whose eigenvalues have negative real parts, but $B + B'$ is not negative definite.
7. (*Nonlinear drift*) (a) Show that if the linear drift is replaced by an arbitrary Lipschitz drift $b(\cdot)$ then Theorem 18.1 holds with no essential change of proof other than replacing $B(x-y)$ in (18.7) and Bx in (18.8) by $b(x) - b(y)$ and $b(x)$, respectively. (b) Verify the stability assertion in Example 7.
8. Consider the case $k = 2$, $b(x) = (b_1(x), b_2(x))$, where $b_1(x) = -cx_1 - \alpha x_1 x_2^2$, $b_2(x) = -dx_2 - \beta x_1^2 x_2$, and c, d, α, β are positive constants. (a) Show that the eigenvalues of $\frac{1}{2}(J(x) + J'(x))$ are no more than $\max\{-c, -d\}$. [*Hint*: Solve the equation $\det(\frac{J(x)+J'(x)}{2} - \lambda) = 0$.] (b) What conclusions can be drawn about stability in distribution in the cases (i) $\sigma(\cdot)$ is constant ? (ii) $\sigma(\cdot)$ is Lipschitz ?

Chapter 19
Stochastic Integrals with L^2-Martingales

In this chapter the theory of stochastic integrals is extended to that with respect to square integrable martingales. This extension is enabled by the so called Doob–Meyer decomposition of submartingales

In this chapter the theory of stochastic integrals is extended to that with respect to square integrable martingales. This extension is enabled by the so called Doob–Meyer decomposition of submartingales, Theorem 19.1, below. In the case of a discrete parameter submartingale $\{X_n : n \geq 0\}$ adapted to a filtration $\{\mathcal{F}_n : n \geq 0\}$, such a decomposition is simple to derive (the Doob decomposition) as follows. Let $Z_n = \mathbb{E}(X_n - X_{n-1}|\mathcal{F}_{n-1}), n \geq 1$, and define $A_n = \sum_{1 \leq m \leq n} Z_m (n \geq 1)$, $A_0 = 0$. Then

$$X_n = M_n + A_n \ (n \geq 0); \quad M_0 = X_0, \ M_n = X_n - A_n. \tag{19.1}$$

Here $A_n(n \geq 0)$ is an increasing nonnegative process, and $M_n(n \geq 0)$ is a martingale, both adapted to the filtration $\{\mathcal{F}_n : n \geq 0\}$. Moreover, A_n $(n \geq 0)$ is required to be predictable, i.e., $\mathcal{F}_{n-1}$-measurable $(n \geq 1)$. The continuous parameter case is more delicate, as we shall see below. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $\{\mathcal{F}_t : t \geq 0\}$ a P-complete, right-continuous filtration $(\mathcal{F}_t \subset \mathcal{F} \ \forall t)$. Consider a right-continuous $\{\mathcal{F}_t\}$-adapted submartingale $\{X_t : t \geq 0\}$, with left limits. Recall that every stochastically continuous submartingale has such a version (Bhattacharya and Waymire (2021), Theorem 13.6). Throughout this chapter, unless otherwise specified, we will assume this set up.

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_19

Definition 19.1 An $\{\mathcal{F}_t\}$-adapted process $\{A_t : t \geq 0\}$ is said to be an integrable increasing process, if (i) $A_0 = 0$ and A_t is increasing with t, a.s., (ii) $t \to A_t$ is a.s. right-continuous, and (iii) $\mathbb{E}(A_t) < \infty$ for all t. Such a process is said to be natural if for every bounded right-continuous $\{\mathcal{F}_t\}$-martingale $\{m_t\}$, one has

$$\mathbb{E}\int_{[0,t]} m_s dA_s = \mathbb{E}\int_{[0,t]} m_{s^-} dA_s, \tag{19.2}$$

or, equivalently,

$$\mathbb{E}(m_t A_t) = \mathbb{E}\int_{[0,t]} m_{s^-} dA_s, \text{ for all } t \geq 0. \tag{19.3}$$

In (19.3), the integration is in the usual sense of Lebesgue–Stieltjes integration (see BCPT[1] p. 228) with respect to the measure dA_s. To see the equivalence of the two conditions in (19.2) and (19.3), consider a partition $\Pi_n : 0 = t_0 < t_1 < \cdots < t_n = t$, of $[0, t]$ with $\Delta_n := \max\{t_{i+1} - t_i : i = 0, \ldots, n-1\} \to 0$ as $n \uparrow \infty$. Then $\mathbb{E}(m_t A_t) = \mathbb{E}(\sum_{0 \leq i \leq n-1} m_t(A_{t_{i+1}} - A_{t_i})) = \mathbb{E}(\sum_{0 \leq i \leq n-1} m_{t_{i+1}}(A_{t_{i+1}} - A_{t_i}))$, using $\mathbb{E}(m_t|\mathcal{F}_{t_{i+1}}) = m_{t_{i+1}} (i = 0, 1, \ldots, n-1)$. Letting $n \uparrow \infty$, and using the right-continuity of m_s, $0 \leq s \leq t$ the last sum converges to $\int_{[0,t]} m_s dA_s$, establishing the equivalence.

Remark 19.1 If an integrable increasing process $\{A_t : t \geq 0\}$ is continuous, almost surely, then it is automatically natural. For on $[0, t]$ there are at most countably many discontinuities of a right-continuous martingale $m_s(\omega)$, for each $\omega \in \Omega$, and with probability one, $dA_s(\omega)$, $0 \leq s \leq t$, has no atom.

Definition 19.2 For each $a > 0$, denote by $\mathcal{T}_a$, the set of all $\{\mathcal{F}_t\}$-stopping times $\tau \leq a$, a.s. A submartingale $\{X_t : t \geq 0\}$ is said to be of class DL, if for every $a > 0$, the family of random variables $\{X_\tau : \tau \in \mathcal{T}_a\}$ is uniformly integrable.

Remark 19.2 Note that every martingale, or nonnegative submartingale, $\{M_t : t \geq 0\}$ is of class DL. For, in view of the optional stopping theorem[2], if $\tau \leq a$, then

$$\int_{[|M_\tau| \geq \lambda]} |M_\tau| dP \leq \int_{[|M_\tau| \geq \lambda]} |M_a| dP \text{ and } P(|M_\tau| \geq \lambda) \leq \mathbb{E}|M_\tau|/\lambda \leq \mathbb{E}|M_a|/\lambda.$$

Theorem 19.1 (Doob–Meyer Decomposition) [3] *Let $\{X_t : t \geq 0\}$, be a right-continuous submartingale of class DL. Then (a) $\{X_t\}$ can be expressed as*

[1] Throughout, BCPT refers to Bhattacharya and Waymire (2016), A Basic Course in Probability Theory.

[2] See BCPT, Theorem 3.8 and Remark 3.7.

[3] The result is due to Meyer (1966). The method of proof is due to Rao (1969), as presented in Ikeda and Watanabe (1989).

$$X_t = M_t + A_t,$$

where $M = \{M_t : t \geq 0\}$ *is a martingale and* $A = \{A_t : t \geq 0\}$ *is an integrable increasing process, which can be chosen to be natural to make this decomposition unique.*

Proof Let us first prove the uniqueness of the decomposition, when A is natural. Suppose $X_t = M_t + A_t = M'_t + A'_t$ are two decompositions. Consider a bounded right-continuous martingale $\{m_s : s \geq 0\}$, with left limits. Fix $t > 0$, and consider partitions $\Pi_n : 0 = t_0 < t_1 < \cdots < t_n = t$, of $[0, t]$ with $\Delta_n := \max\{t_{i+1} - t_i : i = 0, \ldots, n-1\} \downarrow 0$ as $n \uparrow \infty$. Then using $A_s - A'_s = M'_s - M_s$ and denoting the martingale $M'_s - M_s = N_s$,

$$\begin{aligned}\mathbb{E}\int_{[0,t]} m_{s-} d(A_s - A'_s) &= \lim_{n\to\infty} \sum_{0\leq i\leq n-1} \mathbb{E}(m_{t_i}[(A_{t_{i+1}} - A'_{t_{i+1}}) - (A_{t_i} - A'_{t_i})]) \\ &= \lim_{n\to\infty} \sum_{0\leq i\leq n-1} \mathbb{E}(m_{t_i}(N_{t_{i+1}} - N_{t_i})) = 0, \end{aligned} \tag{19.4}$$

since $\mathbb{E}(N_{t_{i+1}}|\mathcal{F}_{t_i}) = N_{t_i}$. Hence we arrive at

$$\mathbb{E}\int_{[0,t]} m_{s-} d(A_s) = \mathbb{E}\int_{[0,t]} m_{s-} d(A'_s).$$

Using the property that A and A' are natural (see (19.3)), one then has $\mathbb{E}(m_t A_t) = \mathbb{E}(m_t A'_t)$ for all bounded right-continuous martingales $\{m_t\}$. In particular, letting Y be any bounded random variable and letting $m_t = \mathbb{E}(Y|\mathcal{F}_t)$, one has

$$\mathbb{E}(YA_t) = \mathbb{E}(m_t A_t) = \mathbb{E}(m_t A'_t) = \mathbb{E}(YA'_t),$$

for any bounded random variable Y. This implies $A_t = A'_t$ almost surely. Since this holds with probability one (simultaneously) on a countable dense subset of $[0, t]$ and A and A' are right-continuous, one gets $A = A'$ a.s. and, therefore $M = M'$ a.s. . Next, let $Y_t = X_t - \mathbb{E}(X_a|\mathcal{F}_t)$, which is a non-positive submartingale on $[0, a]$, $Y_a = 0$ a.s.. Consider the partitions $\Pi_n : t_{j,n} = ja/2^n : j = 0, 1, .., 2^n\}(n = 1, 2, \ldots)$, of $[0, a]$. Then, as in the Doob decomposition (19.1),

$$Y_t = A_t^{(n)} + m(n)_t, \quad m(n)_t = Y_t - A_t^{(n)}, \tag{19.5}$$

where $m(n)_t = \mathbb{E}(m(n)_a|\mathcal{F}_t),\ t \in \Pi_n$, (with $t = t_{k,n} = ka/2^n$), is a martingale. Here

$$A_{t_{k,n}}^{(n)} = \sum_{0\leq j\leq k-1} [\mathbb{E}(Y_{t_{j+1,n}}|\mathcal{F}_{t_{j,n}}) - Y_{t_{j,n}}]\ (k = 0, 1, \ldots, 2^n),$$

is, for each n, an increasing process $A^{(n)}$. We will first show that

$$\{A_a^{(n)} : n = 1, 2, \dots\} \text{ is uniformly integrable.} \tag{19.6}$$

For this, let

$$\tau(n)_\lambda = \inf_k \{t_{k-1,n} : A_{t_{k,n}}^{(n)} > \lambda\},$$

if the infimum is attained, or $\tau(n)_\lambda = a$, otherwise. Then $\tau(n)_\lambda \in \mathcal{T}_a$. Note that, for $t \in \Pi_n, Y_t = -\mathbb{E}(A_a^{(n)}|\mathcal{F}_t) + A_t^{(n)}$, since $m(n)_t = \mathbb{E}(m(n)_a|\mathcal{F}_t) = \mathbb{E}(Y_a - A_a^{(n)}|\mathcal{F}_t) = -\mathbb{E}(A_a^{(n)}|\mathcal{F}_t)$. By the optional sampling theorem, $Y_{t\wedge\tau(n)_\lambda} = -\mathbb{E}(A_a^{(n)}|\mathcal{F}_{t\wedge\tau(n)_\lambda}) + A_{t\wedge\tau(n)_\lambda}^{(n)}$ $(t \in \Pi_n)$ is a submartingale. Also, letting $t = \tau(n)_\lambda$, one has

$$\begin{aligned} Y_{\tau(n)_\lambda} &= -\mathbb{E}(A_a^{(n)}|\mathcal{F}_{\tau(n)_\lambda}) + A_{\tau(n)_\lambda}^{(n)} \\ &\leq \lambda - \mathbb{E}(A_a^{(n)}|\mathcal{F}_{\tau(n)_\lambda}), \end{aligned} \tag{19.7}$$

noting that $A_{\tau(n)_\lambda}^{(n)} \leq \lambda$ (by definition of $\tau(n)_\lambda$). Also, $[\tau(n)_\lambda < a] = [A_a^{(n)} > \lambda]$. Hence, taking expectations in (19.7), on the event $[\tau(n)_\lambda < a] = [A_a^{(n)} > \lambda]$, to get

$$\mathbb{E}(A_a^{(n)}\mathbf{1}_{[A_a^{(n)}>\lambda]}) \leq -\mathbb{E}(Y_{\tau(n)_\lambda}\mathbf{1}_{[A_a^{(n)}>\lambda]}) + \lambda P(\tau(n)_\lambda < a). \tag{19.8}$$

One may use the equality in (19.7) with $\frac{\lambda}{2}$ in place of λ and taking expectations on the event $[\tau(n)_{\frac{\lambda}{2}} < a]$ to get

$$\begin{aligned} -\mathbb{E}(Y_{\tau(n)_{\frac{\lambda}{2}}}\mathbf{1}_{[\tau(n)_{\lambda/2}<a]}) &= \mathbb{E}((A_a^{(n)} - A_{\tau(n)_{\frac{\lambda}{2}}}^{(n)})\mathbf{1}_{[\tau(n)_{\frac{\lambda}{2}}<a]}) \\ &\geq \mathbb{E}((A_a^{(n)} - A_{\tau(n)_{\frac{\lambda}{2}}}^{(n)})\mathbf{1}_{[\tau(n)_\lambda<a]}) \\ &\geq (\lambda - \frac{\lambda}{2})P(\tau(n)_\lambda < a) \\ &= \frac{\lambda}{2}(P(\tau(n)_\lambda < a), \end{aligned}$$

one has

$$\lambda P(\tau(n)_\lambda < a) \leq -2\mathbb{E}(Y_{\tau(n)_{\frac{\lambda}{2}}}\mathbf{1}_{[\tau(n)_{\frac{\lambda}{2}}<a]}). \tag{19.9}$$

Comparing (19.8) and (19.9), one obtains

$$\begin{aligned}
&\mathbb{E}((A^{(n)}a)\mathbf{1}_{[A_a^{(n)}>\lambda]}) \\
&\quad\le -\mathbb{E}(Y_{\tau(n)_\lambda}\mathbf{1}_{[\tau(n)_\lambda<a]}) - 2\mathbb{E}(Y_{\tau(n)_{\frac{\lambda}{2}}}\mathbf{1}_{[\tau(n)_{\frac{\lambda}{2}}<a]}) \\
&\quad\le \mathbb{E}X_{\tau(n)_\lambda}\mathbf{1}_{[\tau(n)_\lambda<a]}) + 2\mathbb{E}(X_{\frac{\tau(n)_\lambda}{2}}\mathbf{1}_{[\tau(n)_{\frac{\lambda}{2}}<a]}) + 2\mathbb{E}|X_a|.
\end{aligned} \tag{19.10}$$

Since $\tau(n)_\lambda, \tau(n)_{\frac{\lambda}{2}} \in \mathcal{T}_a$ and X belongs to the class DL, the proof of uniform integrability of $\{A_a^{(n)} : n = 0, 1, \ldots, 2^n\}$ is complete, since $P(\tau(n)_{\frac{\lambda}{2}} < a) = P(A_a^{(n)} > \frac{\lambda}{2}) \ge \mathbb{E}(A_a^{(n)})/(\lambda/2)$. Note that from from (19.5), one gets $\mathbb{E}(A_a^{(n)}) \le \mathbb{E}(-Y_0)$. Now from (19.11) one concludes that $\{A_a^{(n)}\}$ is uniformly integrable, and, therefore, by the Dunford–Pettis compactness criterion (Lemma 1 below), there exists a subsequence of $\{A(n')_a\}$ which converges to some integrable A_a in the weak topology $\sigma(L^1, L^\infty)$, i.e., for every bounded random variable Z, $\mathbb{E}(A(n')_a Z) \to \mathbb{E}(A_a Z)$ as $n' \to \infty$. Note that, by the same argument, there exists a subsequence of $\{A_t^{(n)}\}$ which converges to some A_t for every t belonging to the countable set $\Pi = \cup_k \Pi_k$. Thus $\{A_t : t \in \Pi\}$ is an integrable increasing $\{\mathcal{F}_t\}$-adapted process indexed by Π. Recall that $Y_t = -\mathbb{E}(A_a^{(n)}|\mathcal{F}_t) + A_t^{(n)}$, $t \in \Pi_n$. Since A_t, so defined, is right-continuous on Π (recall that $\{\mathcal{F}_t\}$ is right-continuous), for $t \in [0, a]$ one can choose a sequence $t_k \downarrow t$ as $k \uparrow \infty$, and using $Y_{t_k} = -\mathbb{E}(A_a^{(n)}|\mathcal{F}_{t_k}) + A_{t_k}^{(n)}$, ($n$ satisfying $t_{k,n} = ka/2^n$), one obtains $A_t = Y_t + \mathbb{E}(A_a|\mathcal{F}_t)$. Hence X has the desired decomposition

$$\begin{aligned}
X_t &= Y_t + \mathbb{E}(X_a|\mathcal{F}_t) \\
&= A_t + \mathbb{E}(X_a - A_a|\mathcal{F}_t) = A_t + M_t.
\end{aligned}$$

It remains to prove that $\{A_t\}$ is natural. Let $\{m_t\}$ be a bounded right-continuous $\{\mathcal{F}_t\}$-martingale on $[0, a]$. Then, summing over $k = 0, 1, \ldots, 2^n - 1$,

$$\begin{aligned}
\mathbb{E}(m_a A_a^{(n)}) &= \mathbb{E}\sum_k (m_a(A_{t_{k+1,n}}^{(n)} - A_{t_{k,n}}^{(n)})) \\
&= \mathbb{E}\sum_k (m_{t_{k,n}}(A_{t_{k+1,n}}^{(n)} - A_{t_{k,n}}^{(n)})) \\
&= \mathbb{E}\sum_k (m_{t_{k,n}}(Y_{t_{k+1,n}} - Y_{t_{k,n}})) \\
&= \mathbb{E}\sum_k (m_{t_{k,n}}(A_{t_{k+1,n}} - A_{t_{k,n}}) \\
&\quad + \mathbb{E}\sum_k m_{t_{k,n}}(\mathbb{E}(A_a|\mathcal{F}_{t_{k+1,n}}) - \mathbb{E}(A_a|\mathcal{F}_{t_{k,n}})).
\end{aligned} \tag{19.11}$$

For the second equality, note that $A^{(n)}_{t_{k+1,n}}$ is $\mathcal{F}_{t_{k,n}}$-measurable, so that one may take conditional expectation, given $\mathcal{F}_{t_{k,n}}$. For the third equality, we used the equality $Y_t = A_t - \mathbb{E}(A_a|\mathcal{F}_t)$. Now the last sum in (19.11) vanishes on taking conditional expectation, given $\mathcal{F}_{t_{k,n}}$, so that one has

$$\mathbb{E}(m_a A^{(n)}_a) = \mathbb{E}\sum_k m_{t_{k,n}}(A_{t_{k+1,n}} - A_{t_{k,n}}).$$

Taking limit as $n \to \infty$, one gets the desired relation

$$\mathbb{E}(m_a A_a) = \mathbb{E}\int_{[0,a]} m_{s-} dA_s. \tag{19.12}$$

■

Lemma 1 (Dunford–Pettis Theorem) *Let $(S, \mathcal{S}, Q)$ be a probability space. A bounded subset D of $L^1(S, \mathcal{S}, Q) \equiv L^1(Q)$ is relatively compact in the weak topology $\sigma(L^1, L^\infty)$ if and only if D is uniformly integrable.*

Proof [4] Suppose the set D in $L^1(Q)$ is uniformly integrable. Consider the map

$$T(f)g = \int_S fg dQ, \quad f \in L^1(Q),\ g \in L^\infty(Q). \tag{19.13}$$

For each $f \in D$, it is a map into $L^\infty(Q)^*$, the dual Banach space of continuous linear functionals on $L^\infty(Q)$. Indeed, the map $f \to Tf$ is an isometry: $||(Tf)^*|| \equiv \sup\{||Tf(g)|| : ||g||_\infty \leq 1\} = ||f||_1$, between D and $T(D)$. Hence $T(D)$ is contained in a bounded set B of $L^\infty(Q)^*$ (bounded by $\sup\{||f||_1 : f \in D\}$) and is, therefore, relatively weak*-compact, $\sigma(L^{\infty *}, L^\infty)$, by Alaoglu's Theorem (See Lemma 2). Thus, given any net $\{f_\beta\} \subset D$, there exists a subnet[5] $\{f_{\beta'}\}$ such that $T(f_{\beta'})$ converges to some element F of B in the $\sigma(L^{\infty *}, L^\infty)$. By the isometry above, using the fact $L^1 \subset L^{\infty **}$, one has $\int_S f_{\beta'} g dQ \to F(g)$. We will show that F is (i.e., can be identified with) an element of L^1. By uniform integrability, for any given $\varepsilon > 0$, there exists $\delta > 0$ such that $\int_A f_{\beta'} < \delta$ if $Q(A) < \varepsilon$ $(A \in \mathcal{S})$. The function $\mu(A) := F(\mathbf{1}_A)$ is clearly finitely additive on $\mathcal{S}$. The uniform integrability condition above now shows that if $A_n \downarrow \emptyset$, $(A_n \in \mathcal{S})$, then $\mu(A_n) \downarrow 0$, and this proves the countable additivity of μ (Exercise 1). Also if $Q(A) = 0$ then

[4] Since we only need the "sufficiency"("if") part of this result as a lemma, we restrict the proof accordingly. See (Dunford and Schwartz, 1963, 1988, p. 294), or Bell (2015) for proofs of both necessary and sufficient conditions.

[5] The concept of net and subnet refers to a generalization of sequence and subsequence, respectively, in topology. For a formal definition see Folland (1984), p. 119.

$\mu(A) = 0$, and by the Radon–Nikodym theorem[6] $d\mu = f dQ$ for some $f \in L^1(Q)$, $F(g) = \int_S fgdQ \; \forall \, g \in L^\infty$. Hence, $f_{\beta'} \to f \in \sigma(L^1, L^\infty)$. ■

For the next result, recall

Definition 19.3 The weak star topology on the dual X^* of a Banach space X is the weakest topology for which the linear functional $\lambda \to \lambda(x) \in \mathbb{R}, \lambda \in X^*$, is continuous for every $x \in X$.

Lemma 2 (Alaoglu Theorem) *A bounded subset of the dual X^* of a Banach space X is relatively compact in the weak star topology: $\sigma(X^*, X)$.*

Proof It is enough to show that the closed unit ball $B = \{\lambda \in X^* : ||\lambda||^* \le 1\}$ of X^* is compact in the $\sigma(X^*, X)$ topology. Note that $B = \{\lambda \in X^* : |\lambda(x)| \le ||x|| \; \forall \, x \in X\}$ can be identified with a subset of $D = \prod_{x \in X}\{z \in \mathbb{R} : |z| \le ||x||\}$, which is the set of all real-valued functions f on X such that $|f(x)| \le ||x||$. By Tychonoff's theorem[7], D is compact in the topology of pointwise convergence. The subset B of D comprises all real-valued linear functions λ on X which are continuous and satisfy $|\lambda(x)| \le ||x|| \; \forall \, x \in X$. We need to show that this subset is closed, when endowed with the topology of pointwise convergence. Given a net $\{\lambda_\beta\} \subset B$, by compactness of D, there exists a subnet $\{\lambda_{\beta'}\}$ such that $\lambda_{\beta'}(x) \to \lambda(x) \; \forall \, x \in X$, for some function $\lambda(x)$. Since $\lambda_{\beta'}(ax + by) = a\lambda_{\beta'}(x) + b\lambda_{\beta'}(y) \; \forall \, a, b \in \mathbb{R}$, pointwise convergence implies $\lambda(ax + by) = a\lambda(x) + b\lambda(y)$ and, clearly, $|\lambda(x)| \le ||x|| \; \forall \, x \in X$, it follows that $\lambda \in B$. Finally, note that the pointwise convergence topology on D when relativized on B is precisely the weak star topology on X^*. ■

We now show that if the submartingale in Theorem 19.1 is almost surely continuous, then so are the two components A and M.

Definition 19.4 A submartingale $X = \{X_t : t \ge 0\}$ is regular if for every $a > 0$,

$$\lim_{n\to\infty} \mathbb{E}X_{\tau_n} = \mathbb{E}X_\tau \; \forall \, \tau_n \uparrow \tau \;\; (\tau_n, \tau \in \mathcal{T}_a). \tag{19.14}$$

Proposition 19.2 [8] *If $\{X_t : t \ge 0\}$ is a submartingale of class DL with the Doob–Meyer decomposition $X = A + M$ (with A natural), then A is continuous if and only if X is regular.*

Proof If A is continuous and $\tau_n \uparrow \tau$, then

$$X_{\tau_n} = A_{\tau_n} + M_{\tau_n} \to A_\tau + M_\tau \; (\tau_n, \tau \in \mathcal{T}_a),$$

[6] See BCPT p. 250.

[7] See BCPT p. 243.

[8] Ikeda and Watanabe (1989), pp. 38–40.

and the expectations converge ($A_{\tau_n} \uparrow A_\tau$ integrable, as $\tau \le a$, and $\mathbb{E}M_{\tau_n} = \mathbb{E}M_\tau$). For the converse, assume that X is regular. Then A is regular, i.e., $\mathbb{E}A_{\tau_n} \uparrow \mathbb{E}A_\tau$. Consider the increasing partitions $\Pi_n = \{t_{j,n} = ja/2^n : j = 0, 1, \ldots, 2^n\}(n = 1, 2, \ldots)$, of $[0, a]$, and define, for $a, \lambda > 0$,

$$A_t^{(n)} = \mathbb{E}(A_{t_{k+1,n}} \wedge \lambda|\mathcal{F}_t) \quad t \in [t_{k,n}, t_{k+1,n}), (k = 0, 1, \ldots, 2^n - 1). \tag{19.15}$$

Then

$$\mathbb{E}\int_{[0,t]} A_{s-}^{(n)} dA_s = \mathbb{E}\int_{[0,t]} A_s^{(n)} dA_s \quad \forall\, t \in [0, a]. \tag{19.16}$$

This uses the fact $\{A^{(n)}\}$ is a right-continuous bounded martingale on each of the subintervals and the equality holds on it, since A is natural. To show that there exists a subsequence $\{n(j) : j = 1, 2, ..\}$ such that $A_t^{(n(j))} \to A_t \wedge \lambda$ uniformly on $[0, a]$ as $j \to \infty$, define for each $\varepsilon > 0$, the stopping times

$$\tau_{n,\varepsilon} = \inf\{t \in [0, a] : A_t^{(n)} - A_t \wedge \lambda > \varepsilon\},$$

where the infimum over the empty set is ∞. Note that for each $n \ge 1$,

$$A_a^{(n)} = \mathbb{E}(A_a \wedge \lambda|\mathcal{F}_a) = A_a \wedge \lambda.$$

Hence $\tau_{n,\varepsilon} = a$ implies

$$A_t^{(n)} - A_t \wedge \lambda \le \varepsilon\ \forall t\ \in [0, a].$$

Let $\Theta_n(t) = t_{k+1,n}$ if $t \in (t_{k,n}, t_{k+1,n}$. Then $\Theta_n(\tau_{n,\varepsilon})$ is a stopping time, $\tau_{n,\varepsilon}$ and $\Theta_n(\tau_{n,\varepsilon}) \in \mathcal{T}_a\ \forall n$. As $A_t^{(n)}$ is decreasing with n, $\tau_{n,\varepsilon}$ is increasing with n; also, $\Theta_n(t) \downarrow t$ as $n \uparrow \infty$, $\Theta_n(\tau_{n,\varepsilon}) \ge \tau_{n,\varepsilon}$. Hence

$$\tau_\varepsilon := \lim_n \tau_{n,\varepsilon} = \lim_n \Theta_n(\tau_{n,\varepsilon}),$$

as $n \to \infty$. Next, on $t_{k,n} \le t < t_{k+1,n}$, the pair $\{A_{\tau_{n,\varepsilon}}^{(n)} := \mathbb{E}(A_{t_{k,n}} \wedge \lambda|\mathcal{F}_{\tau_{n,\varepsilon}}), A_{\Theta_n(\tau_{n,\varepsilon})}^{(n)} := \mathbb{E}(A_{t_{k,n}} \wedge \lambda|\mathcal{F}_{\Theta_n(\tau_{n,\varepsilon})}) = A_{\Theta_n(\tau_{n,\varepsilon})} \wedge \lambda)\}$ is a martingale. Therefore,

$$\mathbb{E}(\mathbf{1}_{[t_{k-1,n} \le \tau_{n,\varepsilon} < t_{k,n}]} A_{\tau_{n,\varepsilon}}^{(n)}) = \mathbb{E}(\mathbf{1}_{[t_{k-1,n} \le \tau_{n,\varepsilon} < t_{k,n}]} A_{\Theta_n(\tau_{n,\varepsilon})} \wedge \lambda). \tag{19.17}$$

Summing over $k = 1, \ldots, 2^n$, one gets

$$\mathbb{E}(A_{\tau_{n,\varepsilon}}^{(n)}) = \mathbb{E}(A_{\Theta_n(\tau_{n,\varepsilon})} \wedge \lambda). \tag{19.18}$$

By (19.18) and recalling that $\tau_{n,\varepsilon} < a$ implies $A(n)_{\tau_{n,\varepsilon}} - A_{\tau_{n,\varepsilon}} \wedge \lambda > \varepsilon$, we have

$$\begin{aligned} \mathbb{E}(A_{\Theta_n(\tau_{n,\varepsilon})} \wedge \lambda - A_{\tau_{n,\varepsilon}} \wedge \lambda) &= \mathbb{E}(A^{(n)}_{\tau_{n,\varepsilon}} - A_{\tau_{n,\varepsilon}} \wedge \lambda) \\ &\geq \varepsilon P(\tau_{n,\varepsilon} < a). \end{aligned} \tag{19.19}$$

Since A is regular, and $\Theta_n(\tau_{n,\varepsilon})$ and $\tau_{n,\varepsilon}$ both converge to τ_ε, the left side of (19.19) goes to zero as $n \to \infty$, implying $P(\tau_{n,\varepsilon} < a) \to 0$ as $n \to \infty$. Therefore, $\lim_n P(\sup_{t\in[0,a]} |A^{(n)}_t - A_t \wedge \lambda| > \varepsilon) \to 0$ as $n \to \infty$. This convergence in probability implies that there exists a subsequence $\{n(j) : j = 1, 2, \ldots\}$ such that $\sup_{t\in[0,a]} |A^{(n(j))}_t - A_t \wedge \lambda| \to 0$ almost surely as $j \to \infty$. Applying this to (19.16), one obtains

$$\mathbb{E} \int_{[0,t]} (A_{s-} \wedge \lambda) dA_s = \mathbb{E} \int_{[0,t]} (A_s \wedge \lambda) dA_s \ \forall\, t \in [0, a], \tag{19.20}$$

which leads to

$$\mathbb{E}\{\int_{[0,t]} (A_s \wedge \lambda - A_{s-} \wedge \lambda) dA_s\} = 0. \tag{19.21}$$

Since the integrand is nonnegative and is positive only at points of discontinuity of A_s, which is at most a countable set, dA_s assigns zero probability at possible points of discontinuity. Thus $s \to A_s \wedge \lambda$ is continuous almost surely. This being true for all λ, the proof is complete that $s \to A_s$ is almost surely continuous. ■

We next turn to the definition of integrals with respect to square integrable martingales, which is similar to those with Brownian motion, replacing independent increments by increments orthogonal to the past. Continue to assume that the underlying probability space $(\Omega, \mathcal{F}, P)$ is P-complete and the σ-fields of the filtration $\{\mathcal{F}_t\}$ are also all P-complete. In addition, assume that $\mathcal{F}_t$ is right-continuous, i.e., $\mathcal{F}_t = \mathcal{F}_{t+} \equiv \cap_{s>t}\mathcal{F}_s$ for every t. Note that every right-continuous submartingale/martingale with respect to a filtration $\{\mathcal{G}_t\}$ is a submartingale/martingale with respect to the filtration $\{\mathcal{G}_{t+}\}$ (Exercise 7). We will then assume that all submartingales/martingales are right-continuous with left limit. The following definitions are preparatory for the definition of stochastic integrals with respect L^2-martingales.

Definition 19.5 We will denote by $\mathcal{M}_2$ the class of square integrable $\{\mathcal{F}_t : t \geq 0\}$ martingales $\{M_t : t \geq 0\}$, right-continuous with left limits. The subclass of continuous square integrable martingales will be denoted by $\mathcal{M}^c_2$. If the process $\{M_t : t \geq 0\}$ is such that there exists a sequence of $\{\mathcal{F}_t\}$-stopping times $\tau_n \uparrow \infty$ almost surely as $n \uparrow \infty$, $\{M_t \wedge \tau_n : t \geq 0\}$ is a square integrable $\{\mathcal{F}_t\}$-martingale $(n = 1, 2, \ldots)$, then $\{M_t : t \geq 0\}$ is said to be a $\{\mathcal{F}_t\}$-local martingale. The class of such local martingales will be denoted by $\mathcal{M}_{2,l}$; the subclass of continuous square integrable $\{\mathcal{F}_t\}$-local martingales will be denoted by $M^c_{2,l}$.

Definition 19.6 A real-valued stochastic process $\{f(t) : t \geq 0\}$ is said to be a non-anticipative step functional (with respect to $\{\mathcal{F}_t\}$) if there exist a sequence $0 = t_0 < t_1 < \cdots < t_m < \cdots$, $t_n \uparrow \infty$ and a sequence of $\mathcal{F}_{t_j}$-measurable random variables $f_j (j = 0, 1, 2, \ldots)$, uniformly bounded by a constant, such that

$$f(t) = f_{j-1} \text{ for } t_{j-1} \leq t < t_j (j = 1, 2, \ldots). \tag{19.22}$$

The class of such functionals is denoted by $\mathcal{L}_0$.

Let $\mathcal{L}_2$ be the class of all $\{\mathcal{F}_t\}$-non-anticipative progressively measurable processes $f(\cdot)$ such that

$$||f||_{2,T}^2 := \mathbb{E} \int_{[0,T]} f^2(s) ds < \infty \quad \forall T > 0.$$

A norm $|| \cdot ||_2$ on $\mathcal{L}_2$ is defined by

$$||f||_2^2 = \sum_{1 \leq k < \infty} 2^{-k} (||f||_2^{2k} \wedge 1). \tag{19.23}$$

When localized to an interval $[\alpha, \beta]$ we will write the processes as $\mathcal{L}_0[\alpha, \beta]$, $\mathcal{L}_2[\alpha, \beta]$, and so on.

Proposition 19.3 *The class $\mathcal{L}_0$ is dense in $\mathcal{L}_2$ in the metric defined by the norm $||f||_2$.*

Proof The proof of this proposition is the same as the proof of the Proposition 6.3 in Chapter 6. ∎

Definition 19.7 Let $M \equiv \{M_t : t \geq 0\} \in \mathcal{M}_2$, and let $M^2 = A + m$ (i.e., $M_t^2 = A_t + m_t$) be the Doob–Meyer decomposition, with A a natural integrable increasing process and m a square integrable martingale. Then $\langle M \rangle := A$, $\langle M, M \rangle_t := A_t$ is said to be the quadratic variation (process) of M. If M and N both belong to $\mathcal{M}_2$, the quadratic co-variation of M and N is defined by $\langle M, N \rangle := \langle (M + N)/2 \rangle - \langle (M - N) \rangle/2 \rangle$. In particular, $\langle M \rangle = \langle M, M \rangle$.

Remark 19.3 It may be shown that $\langle M \rangle_t$ is indeed the quadratic variation of M on $[0, t]$ in the usual classical sense (Exercise 9). The definition below of the stochastic integral of a non-anticipative step functional $f(\cdot)$ defined by (19.22) with respect to an $M \in \mathcal{M}_2^c$ is quite analogous to that with respect to a standard Brownian motion (see Chapter 6, Definition 6.5)).

Definition 19.8 Let f be a non-anticipative step functional defined in (19.22), and $M = \{M_t : t \geq 0\} \in \mathcal{M}_2^c$.

(i) Then the stochastic integral of f (with respect to M) is defined by

$$I_t(f) \equiv M_t = \sum_{0 \le j \le m-1} f_j(M_{t_{j+1}} - M_{t_j}) + f_m(M_t - M_{t_m}),\ t \in [t_m, t_{m+1}), t \ge 0. \tag{19.24}$$

(ii) The stochastic integral of an $f \in \mathcal{L}_2$ is defined to be the limit of a (any) sequence of non-anticipative step functionals $\{f_n\}$ which converges to f uniformly on finite intervals and in the norm (19.24).

Remark 19.4 That the limit in (ii) of Definition 19.8 is well-defined a.s. is proved in the same manner as Proposition 6.2 Theorem 6.5 of Chapter 6 (see Definition 6.5 there).

Theorem 19.4 *Let $M \in \mathcal{M}_2^c$. For $f \in \mathcal{L}_2$. Then the following hold:*

(i) $I_t(0) = 0$ *a.s.*
(ii) $\mathbb{E}(I_t(f) - I_s(f)|\mathcal{F}_s) = 0$ *a.s.* $\forall\, 0 \le s < t$.
(iii) $\mathbb{E}((I_t(f) - I_s(f))^2|\mathcal{F}_s) = \mathbb{E}(\int_{[s,t]} f^2(u)d\langle M\rangle_u|\mathcal{F}_s)$ *a.s.* $(0 \le s < t)$.
(iv) *If $f, g \in \mathcal{L}_2$, then for* $0 \le s < t$,

$$\mathbb{E}((I_t(f) - I_s(f))(I_t(g) - I_s(g))|\mathcal{F}_s) = \mathbb{E}(\int_{[s,t]} f(u)g(u)d\langle M\rangle_u|\mathcal{F}_s)\ a.s..$$

(v) *If $\tau_1 \le \tau_2$ are $\{\mathcal{F}_t\}$-stopping times, then (ii)–(iv) hold with t replaced by $t \wedge \tau_2$ and s replaced by $t \wedge \tau_1$.*

Proof For $f \in \mathcal{L}_0$, the proof is the same as that of Proposition 6.2. For $f \in \mathcal{L}_2$, the proof follows that of Theorem 6.5 of Chapter 6 (Exercise 5). ■

Before concluding this chapter, let us show that (1) the term quadratic variation for a martingale in $\mathcal{M}_2^c$ is appropriate in the usual sense and derive from this the important fact, and (2) if $M \in \mathcal{M}_2^c$ has a finite variation on $[0, a]$, then it is almost surely a constant.

Theorem 19.5 *Let $M \in \mathcal{M}_2^c$, and $\Pi_n : t_{j,n} = ja/2^n, j = 0, 1, \dots, 2^n (n = 1, 2, \dots)$ be partitions of* $[0, a]$, $a > 0$. *(a) Then*

$$\mathbb{E}(\sum_{1 \le j \le 2^n} (M_{t_{j,n}} - M_{t_{j-1,n}})^2 - \langle M_a\rangle)^2 \to 0$$

as $n \to \infty$. In particular,

$$\sum_{1 \le j \le 2^n} (M_{t_{j,n}} - M_{t_{j-1,n}})^2 \to \langle M\rangle_a$$

in probability. (b) $\sum_{1 \le j \le 2^n} |X_{t_{j,n}} - X_{t_{j-1,n}}| \to \infty$ *almost surely as $n \to \infty$.*

Proof [9] (a) First assume that M is bounded, $|M_t| \leq K$ on $[0, a]$. For $0 \leq s < t \leq t_1 < t_2$, one has

$$\begin{aligned}\mathbb{E}((M_{t_2} - M_{t_1})^2|\mathcal{F}_t) &= \mathbb{E}((\mathbb{E}(M_{t_2} - M_{t_1})^2|\mathcal{F}_{t_1})|\mathcal{F}_t) \\ &= \mathbb{E}((M_{t_2}^2 - M_{t_1}^2)|\mathcal{F}_{t_1})|\mathcal{F}_t) = \mathbb{E}(M_{t_2}^2 - M_{t_1}^2)|\mathcal{F}_t),\end{aligned}$$

since $\mathbb{E}(M_{t_2}M_{t_1}|\mathcal{F}_{t_1}) = M_{t_1}^2$ a.s. By the Doob Meyer-decomposition, this may be expressed as

$$\begin{aligned}\mathbb{E}((M_{t_2} - M_{t_1})^2|\mathcal{F}_t) &= \mathbb{E}(M_{t_2}^2 - M_{t_1}^2)|\mathcal{F}_t) \\ &= \mathbb{E}(\langle M\rangle_{t_2} - \langle M\rangle_{t_1}|\mathcal{F}_t)\ a.s. \end{aligned} \tag{19.25}$$

One then has

$$\begin{aligned}&\mathbb{E}((M_{t_2} - M_{t_1})^2|\mathcal{F}_t) \\ &= \mathbb{E}(\sum_{1\leq j\leq 2^n} (M_{t_{j,n}} - M_{t_{j-1,n}})^2 - \langle M\rangle_a)^2 \\ &= \mathbb{E} \sum_{1\leq j\leq 2^n} [(M_{t_{j,n}} - M_{t_{j-1,n}})^2 - (\langle M_{t_{j,n}}\rangle - \langle M_{t_{j-1,n}}\rangle)]^2 \\ &= \sum_{1\leq j\leq 2^n} \mathbb{E}[(M_{t_{j,n}} - M_{t_{j-1,n}})^2 - (\langle M_{t_{j,n}}\rangle - \langle M_{t_{j-1,n}}\rangle)]^2,\end{aligned} \tag{19.26}$$

since the expectation of the product terms vanish by (19.25), i.e., for $k > j$,

$$\begin{aligned}&\mathbb{E}[(M_{t_{j,n}} - M_{t_{j-1,n}})^2 - (\langle M_{t_{j,n}}\rangle - \langle M_{t_{j-1,n}}\rangle] \\ &\quad\times[(M_{t_{k,n}} - M_{t_{j-1,n}})^2 - (\langle M_{t_{k,n}}\rangle - \langle M_{t_{k-1,n}}\rangle)] \\ &= \mathbb{E}\mathbb{E}[(M_{t_{j,n}} - M_{t_{j-1,n}})^2 - (\langle M_{t_{j,n}}\rangle - \langle M_{t_{j-1,n}}\rangle] \\ &\quad\times[(M_{t_{k,n}} - M_{t_{j-1,n}})^2 - (\langle M_{t_{k,n}}\rangle - \langle M_{t_{k-1,n}}\rangle)|\mathcal{F}_{t_{j,n}}],\end{aligned}$$

and

$$\mathbb{E}([(M_{t_{k,n}} - M_{t_{j-1,n}})^2 - (\langle M_{t_{k,n}}\rangle - \langle M_{t_{k-1,n}}\rangle)]\mathcal{F}_{t_{j,n}}) = 0\ a.s.$$

Now the last expression of (19.26) is bounded above by $2(\sum_{1\leq j\leq 2^n} \mathbb{E}(M_{t_{j,n}} - M_{t_{j-1,n}})^4 + 2\mathbb{E}(\sum_{1\leq j\leq 2^n}(\langle M\rangle_{t_{j,n}} - \langle M\rangle_{t_{j-1,n}})^2)$. The expected value of the sum of the first term goes to zero as $n \to \infty$ (Exercise 4(b)). The expectation of the

[9] This proof follows Karatzas and Shreve (1991).

second sum is bounded by $\mathbb{E}\delta_n \langle M \rangle_a$, where $\delta_n = \max\{\langle M \rangle_t - \langle M \rangle_s : 0 \le s < t, t - s \le a2^{-n}\} \downarrow 0$ almost surely as $n \uparrow \infty$. By monotone convergence theorem, the expectation goes to zero. If M is not bounded on $[0, a]$, define the stopping time

$$\tau_K := \inf\{t \ge 0 : |M_t| \ge K \text{ or } \langle M \rangle_t \ge K\},$$

where K is a positive integer. Consider the martingale $M_t^{(K)} := M_t \wedge \tau_K$. It is simple to check that $M^{(K)}$ is an $\{\mathcal{F}_t\}$-martingale (Exercise 6). Note that $M^2_{t\wedge\tau_K} - \langle M \rangle_{t\wedge\tau_K}$ is a bounded martingale. By the uniqueness of the Doob–Meyer decomposition, it follows that $\langle M^{(K)} \rangle_t = \langle M \rangle_{t\wedge\tau_K}$. Hence for partitions Π_n, one has

$$\lim_{n\to\infty} \mathbb{E}[\sum_{1\le j\le 2^n} \left(M^{(K)}_{t_{j,n}} - M^{(K)}_{t_{j-1,n}}\right)^2 - \langle M \rangle_{a\wedge\tau_K}]^2 = 0,$$

appealing to the result for the bounded case. Because $\tau_K \uparrow \infty$ almost surely as $K \uparrow \infty$ and, especially, $P(\tau_K < a) \to 0$ as $K \uparrow \infty$, one obtains the desired convergence in (a). (b) This follows from (a) (see Exercise 3(c)). ■

Remark 19.5 The relation

$$\langle M, N \rangle = \langle (M^2 + N^2 + 2MN)/4 \rangle - \langle (M^2 + N^2 - 2MN)/4 \rangle$$

implies that $\langle M, N \rangle$ may be evaluated as the bounded quadratic variation part of MN expressed uniquely as a function of bounded variation starting at zero, plus a martingale, easing the computation perhaps.

Remark 19.6 For the proof of part (a), it is not necessary to take any special sequence of partitions, as long as the mesh size goes to zero. For the proof of the almost sure result in part (b), any sequence of partitions which become finer with n suffices, since that ensures an increasing sequence of variations, because of the triangle inequality of real numbers.

We now turn to two results which are sometimes useful in different contexts.

Proposition 19.6 (Itô's Formula for Square Integrable Martingales) *Suppose $M \in \mathcal{M}_2^c$ and $f \in C^2(\mathbb{R})$. Then*

$$f(M_t) - f(M_0) = \int_{[0,t]} f'(M_s) dMs + (1/2) \int_{[0,t]} f''(M_s) d\langle M \rangle_s.$$

Proof The proof is basically the same as for Itô's Lemma for stochastic integrals with respect to Brownian motion B_s with the quadratic variation dt replaced by $d\langle M \rangle_t = dA_t$, (Exercise 8). ■

Proposition 19.7 (Maximal Inequality and Quadratic Variation) *Let* $M \in \mathcal{M}_2^c$*, and* $M_t^* = \max\{|M_s| : 0 \le s \le t\}$*,* $M_0 = 0$*. Then for* $p \ge 1$*, one has*

$$\mathbb{E}M_t^{*2p} \le C(p)\mathbb{E}\langle M, M\rangle_t^p,$$

where $C(p)$ *is a constant depending on* p*.*

Proof We will use Doob's maximal inequality for moments, namely,

$$\mathbb{E}M_t^{*2p} \le (2p/(2p-1))^{2p}\mathbb{E}|M_t|^{2p}, \quad p > 1/2. \tag{19.27}$$

For $p = 1$, this yields

$$\mathbb{E}M_t^{*2} \le 4\mathbb{E}|M_t|^2 = 4\mathbb{E}(\langle M, M\rangle_t).$$

For $p > 1$, apply Proposition 19.6 to $f(x) = |x|^{2p}$ to get

$$|M_t|^{2p} = \int_{[0,t]} 2p|M_s|^{2p-1}\mathrm{sgn}(M_s)dM_s + p(2p-1)\int_{[0,t]} |M_s|^{2p-2}d\langle M\rangle_s.$$

Take expectations to get

$$\mathbb{E}|M_t|^{2p} \le p(2p-1)\mathbb{E}|M_t^*|^{2p-2}\langle M\rangle_t) \le p(2p-1)(\mathbb{E}M_t^{*2p})^{1-1/p}(\mathbb{E}\langle M^p\rangle_t)^{1/p}.$$

Now apply (19.27). ■

Exercises

1. Let $(S, \mathcal{S})$ be a measurable space and μ a finitely additive measure on the σ-field $\mathcal{S}$. Suppose that for all $A_n, A \in \mathcal{S}$, $A_n \downarrow \emptyset$, $\mu(A_n) \downarrow 0$. Show that μ is countably additive.
2. Let $\mathbf{B} = \{(B_1(t), B_2(t), B_3(t))\}$ be a standard Brownian motion in $\mathbb{R}^3$, and define $M_t = \frac{1}{R_t}$, $t \ge 1$, where $\{R_t = |\mathbf{B}|_t\}$ is the Bessel process (Chapter 13, Example 6). Show that $\sup_{t\ge1} \mathbb{E}M_t^p < \infty$ for $p < 3$, and that $\{M_t : t \ge 1\}$ is a local martingale, but not a martingale. [*Hint*: R_t^2 is the sum of three i.i.d. Gamma (chi-square) distributed random variables.]
3. Let $\{X_t : t \ge 0\}$ be an almost surely continuous real-valued process, $\Pi_n : t_{j,n} = ja/2^n$, $j = 0, 1, .., 2^n (n = 1, 2, \ldots)$, be partitions of $[0, a]$, $a > 0$. Assume that, for some $p > 0$, $\sum_{1\le j\le 2^n} |X_{t_{j,n}} - X_{t_{j-1,n}}|^p \to Z$ in probability, where Z takes values in $[0, \infty)$. Prove that (a) if $q > p$, then $\sum_{1\le j\le 2^n} |X_{t_{j,n}} - X_{t_{j-1,n}}|^q \to 0$ in probability. [*Hint*: Let $\delta_n := \max\{|X_t - X_s|^{q-p} : 0 \le s < t \le a, t - s \le a2^{-n}\}$. Then the sum is less than $\delta_n \sum_{1\le j\le 2^n} |X_{t_{j,n}} - X_{t_{j-1,n}}|^p$.] (b) Prove that

if $0 < q < p$, then $\sum_{1\le j\le 2^n} |X_{t_{j,n}} - X_{t_{j-1,n}}|^q \to \infty$ in probability on $[Z > 0]$. [*Hint*: Interchange p and q in the above argument, and then divide by δ_n.] (c) Suppose $p > 1$. Prove that the variation defined by $\lim_n \sum_{1\le j\le 2^n} |X_{t_{j,n}} - X_{t_{j-1,n}}| = \infty$ almost surely on $[Z > 0]$. [*Hint*: The sum is monotone increasing in n.]

4. Let $\Pi_n : t_{j,n} = ja/2^n, j = 0, 1, .., 2^n (n = 1, 2, \dots)$ be partitions of $[0, a]$, and let $M \in \mathcal{M}_2^c$, $M_0 = 0$. (a) Assume $|M_t| \le N$ for $N > 0$. Prove that $\mathbb{E}[\sum_{1\le j\le 2^n}(M_{t_{j,n}} - M_{t_{j-1,n}})|^2] \le 6N^4$. [*Hint*: $\mathbb{E}[\sum_{1\le j\le 2^n}(M_{t_{j,n}} - M_{t_{j-1,n}})^2|\mathcal{F}_{t_{j,n}}) = \mathbb{E}[\sum_{1\le j\le 2^n}(M^2_{t_{j,n}} - M^2_{t_{j-1,n}})|\mathcal{F}_{t_{k,n}}] \le \mathbb{E}[X_a^2]|\mathcal{F}_{t_{k,n}}] \le N^2$. Hence, $\mathbb{E}[\sum_{1\le j\le 2^n}\sum_{J+1\le k\le 2^n}(M_{t_{j,n}} - M_{t_{j-1,n}})^2(M_{t_{k,n}} - M_{t_{k-1,n}})^2 = \mathbb{E}[\sum_{1\le j\le 2^n}(M_{t_{j,n}} - M_{t\,j-1,n})^2\mathbb{E}(\sum_{J+1\le k\le 2^n}(M_{t_{k,n}} - M_{t_{k-1},n})^2|\mathcal{F}_{t_{j,n}})] \le N^2N^2 = N^4$. Now use

$$\mathbb{E}[\sum_{1\le j\le 2^n}(M_{t_{j,n}} - M_{t_{j-1,n}})^4 \le 4N^2\mathbb{E}[\sum_{1\le j\le 2^n}(M_{t_{j,n}} - M_{t_{j-1,n}})^2] \le 4N^4.$$

Finally, one has

$$\begin{aligned}
&\mathbb{E}(\sum_{1\le j\le 2^n}(M_{t_{j,n}} - M_{t_{j-1,n}})^2)^2 \\
&= \mathbb{E}\sum_{1\le j\le 2^n}|M_{t_{j,n}} - M_{t_{j-1,n}}|^4 + 2\mathbb{E}[\sum_{1\le j\le 2^n}\sum_{J+1\le k\le 2^n}(M_{t_{j,n}} - M_{t_{j-1,n}})^2 \\
&\qquad\qquad\qquad\qquad \times(M_{t_{k,n}} - M_{t_{k-1,n}})^2 \\
&\le 6N^4.
\end{aligned}$$

(b) Under the hypothesis in (a), show that $\sum_{1\le j\le 2^n}\mathbb{E}(M_{t_{j,n}} - M_{t_{j-1,n}})^4 \to 0$ as $n \to \infty$. [*Hint*: The expectation is bounded by $\mathbb{E}(\delta_n\sum_{1\le j\le 2^n}(M_{t_{j,n}} - M_{t_{j-1,n}})^2)$, where $\delta_n = \max\{(M_t - M_s)^2 : 0 \le s < t, t - s \le a2^{-n}, t \in [0, a]\}$. Hence the expectation goes to 0, noting that $\delta_n \le 4N^2$, $\delta_n \to 0$ as $n \to \infty$, by continuity of M, and using (a).]

5. Prove Theorem 19.4 following the method of proofs of Proposition 6.2 and Theorem 6.5 in Chapter 6.
6. Show that $M^{(K)} = M_{t\wedge\tau_K}$ is an $\{\mathcal{F}_t\}$-martingale.
7. Suppose that $\{X_t : t \ge 0\}$ is a right-continuous martingale/submartingale with respect to a filtration$\{\mathcal{F}_t\}$. Show that $\{X_t : t \ge 0\}$ is a martingale/submartingale with respect to $\{F_{t+}\}$.
8. Write out a proof of Proposition 19.6 following the proof for stochastic integrals with Brownian motion $\{B_s\}$ with M_s replaced by B_s. [*Hint*: Use Theorem 19.5.]
9. Complete the proof of the statement in Remark 19.3.

Chapter 20
Local Time for Brownian Motion

Lévy's local time for Brownian motion may be informally viewed as a (random) 'density' $\ell(t, x)$ for the random measure of time during an interval of time $[0, t]$ at which the Brownian motion occupies an infinitesimally small spatial neighborhood of the point x. An interesting formula due to Skorokhod is used to obtain a formula for the local time at zero. A refined formula due to Tanaka is also included.

Brownian paths are continuous but otherwise quite irregular and chaotic, as is indicated by the fact that when a path hits a point x, it visits every neighborhood of x, however small, infinitely many times, before going off somewhere else. Recall also that the motion is instantaneous, i.e., there is no interval of constancy. This makes Paul Lévy's local time of Brownian motion appear magical! Outside a set of probability zero, a path ω has an *occupation time density* (with respect to Lebesgue measure), or *local time*, of the amount of time in $[0, t]$ that it spends in a set. As we shall indicate in Chapter 21, this local time may be used to construct all of Feller's one-dimensional (instantaneous) diffusions on $\mathbb{R}$.

Theorem 20.1 (Lévy (1948)) *Let $\{B_t : t \geq 0\}$ be a standard Brownian motion defined on a probability space $(\Omega, \mathcal{F}, P)$, starting at some point. Outside a P-null set, there exists a function $\ell(t, x) \geq 0$, continuous in (t, x) and increasing with t, such that*

$$2\int_A \ell(t, x)dx = \int_{[0,t]} \mathbf{1}_{[B_s \in A]}ds \ \forall \text{ Borel sets } A \subset \mathbb{R}, t \geq 0. \tag{20.1}$$

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_20

Definition 20.1 The function $\ell(t,x)$ in Theorem 20.1 is called the *local time* of Brownian motion, and (20.1) is referred to as the *occupation time formula*.

Remark 20.1 The factor 2 on the left side of (20.1) seems anomalous, but it is harmless and commonly used (because Brownian motion is generated by half times the Laplacian and this half factor naturally appears in Itô's Lemma).

Some heuristics are useful for motivating the proof of Theorem 20.1. One may informally think of $2\ell(t,x)$ as a smooth version of $\int_{[0,t]} \delta_x(B_s)ds$, where $\delta_x(\cdot)$ is the Dirac delta function at the point x. If one integrates over x, one then has $\int_A \ell(t,x)dx$ approximately given by

$$\frac{1}{2}\int_A [\int_{[0,t]} \delta_x(B_s)ds]dx = \frac{1}{2}\int_{[0,t]} \mathbf{1}_{[B_s\in A]}ds.$$

Recalling Itô's Lemma, one then looks for a smooth (twice differentiable) function f such that $f''(z)$ is approximately the "density" of $\delta_x(z)$. Such a nice density is $g_n(z;x)$ which is a symmetric p.d.f around x vanishing outside $[x-1/n, x+1/n]$. Then the corresponding $f = f_n$ is, given by

$$f_n(x) = \int_{(-\infty,x]}\int_{(-\infty,z]} g_n(y;x)dydz.$$

By Itô's Lemma,

$$f_n(B_t) - f_n(B_0) = \int_{[0,t]} f'(B_s)dB_s + \frac{1}{2}\int_{[0,t]} f_n''(B_s)ds. \tag{20.2}$$

Thus local time ℓ exists if $1/2\int_{[0,t]} f_n''(B_s)ds \equiv 1/2\int_{[0,t]} g_n(B_s;x)ds$ converges, as $n\to\infty$, to a function satisfying the conditions specified in the theorem. Given the specification of g_n, one has

$$f_n'(y) \to \begin{cases} 1 & \text{if } y > x \\ 1/2 & \text{if } y = x \\ 0 & \text{if } y < x, \end{cases} \tag{20.3}$$

and $f_n(z) \to (z-x)^+$ as $n\to\infty$. From (20.2), (20.3), given the appropriate convergence, $\ell(t,x)$ should be given by

$$\ell(t,x) = (B_t - x)^+ - (B_0 - x)^+ - \int_{[0,t]} \mathbf{1}_{(x,\infty)}(B_s)dB_s. \tag{20.4}$$

Proof *(of Theorem 20.1)* Having[1] arrived to (20.4) by heuristics, we will actually show the random function (20.4) indeed satisfies the requirements of the theorem. To prove this consider, for an arbitrary $T > 0$, the random function $X^x := \{X_t : 0 \le t \le T\}$, where X_t is given by the stochastic integral in (20.4). Then $x \to X^x$ is a random function on $\mathbb{R}$ into the metric space $C([0, T] : \mathbb{R})$ of real-valued continuous functions f on $[0, T]$ endowed with the sup norm:

$$||f|| = \max\{|f(s)| : 0 \le s \le T\}.$$

Then, by Lemma 1 below,

$$\mathbb{E}||X^x - X^z||^4 \le c(T)|x - z|^2. \tag{20.5}$$

for some constant depending only on T. By the Kolmogorov–Chentsov Theorem (see BCPT[2] Theorem 10.2, p. 180), it follows that $x \to X^x$ is almost surely continuous in the sup norm. Thus the random function in (20.4) is continuous on $[0, \infty) \times \mathbb{R}$, almost surely. To prove (20.1), let f be a continuous function on $\mathbb{R}$ with compact support, and

$$\begin{aligned} F(y) &= \int_{(-\infty,\infty)} (y - x)^+ f(x)dx \\ &= \int_{(-\infty,y]} (y - x) f(x)dx = \int_{[0,\infty)} z f(y - z)dz. \end{aligned}$$

Then, integrating by parts,

$$\begin{aligned} F'(y) &= \int_{[0,\infty)} z f'(y - z)dz \\ &= \int_{(-\infty,y]} f(u)du, \end{aligned}$$

and

$$F''(y) = f(y).$$

By Itô's Lemma,

$$F(B_t) - F(B_0) - \int_{[0,t]} F'(B_s)dB_s = 1/2 \int_{[0,t]} f(B_s)ds. \tag{20.6}$$

[1] Ikeda and Watanabe (1989), pp. 113–116.

[2] Throughout, BCPT refers to Bhattacharya and Waymire (2016), A Basic Course in Probability Theory.

The integral on the left equals, by a Fubini type interchange of order (See Lemma 2),

$$\int_{[0,t]} F'(B_s)dB_s = \int_{[0,t]} (\int_{(-\infty,\infty)} f(u)\mathbf{1}_{(u,\infty)}(B_s)du)dB_s.$$

Combining this with the remaining term on the left side of (20.6), one gets

$$\begin{aligned}
&F(B_t) - F(B_0) - \int_{[0,t]} F'(B_s)dB_s \\
&= \int_{(-\infty,\infty)} (f(u)\{(B_t - u)^+ - (B_0 - u)^+\} - \int_{[0,t]} \mathbf{1}_{[u,\infty)}(B_s)dB_s)du \\
&= 1/2 \int_{[0,t]} f(B_s)ds.
\end{aligned}$$

In particular, the expression in (20.4) is indeed the local time $\ell(t, x)$. ■

Lemma 1 *Let* $X^x(t) = \int_{[0,t]} \mathbf{1}_{(x,\infty)}(B_s)dB_s$, *where* $\{B_s : s \geq 0\}$ *is a standard Brownian motion. Then*

$$\mathbb{E}||X^x - X^z|| \equiv \mathbb{E} \sup_{0\leq t\leq T} |X^x(t) - X^z(t)| \leq c(T)|x - z|^2,$$

where $c(T)$ *is a constant depending only on* T.

Proof Use the martingale moment inequality with quadratic variation (Proposition 19.7) for the martingale

$$M_t \equiv X^x(t) - X^z(t) = \int_{[0,t]} \mathbf{1}_{(x,z]}(B_s)dB_s, 0 \leq t \leq T,$$

(for $x > z$, say), to obtain

$$\mathbb{E}||X^x - X^z||^4 \leq c(2)\mathbb{E}\langle M\rangle_T^2, \tag{20.7}$$

where $\langle M\rangle_T$ is the quadratic variation of M on $[0, T]$:

$$\langle M\rangle_T = \int_{[0,T]} \mathbf{1}_{(x,z]}(B_s)ds. \tag{20.8}$$

$$\begin{aligned}
\mathbb{E}\langle M\rangle_T^2 &= \mathbb{E}\{\int_{[0,T]} \mathbf{1}_{(x,z]}(B_s)ds)(\int_{[0,T]} \mathbf{1}_{(x,z]}(B_s)ds)\} \\
&= 2\int_{[0,T]} (\int_{[s,T]} \mathbb{E}\mathbf{1}_{(x,z]}(B_s)\mathbf{1}_{(x,z]}(B_t)dt)ds.
\end{aligned} \tag{20.9}$$

Now for $s < t$, the conditional distribution of B_t, given B_s, is normally distributed with mean B_s and variance $t - s$. Hence the conditional probability that $B_t \in (z, x]$, given B_s, is

$$\Phi((x - B_s)/\sqrt{t-s}) - \Phi((z - B_s)/\sqrt{t-s}) \le c|x - z|/\sqrt{t-s},$$

for some absolute constant c. Also, if $B_0 = x_0$,

$$P(z < B_s \le x) P(z - x_0 \le B_s - x_0 \le x - x_0) \le c|x - z|/\sqrt{s}.$$

Hence

$$\begin{aligned}
\mathbb{E}\langle M\rangle_T^2 &\le 2\int_{[0,T]}(\int_{[s,T]}(t-s)^{-1/2}dt)s^{-1/2}ds \\
&= 2c^2|x-z|^2\int_{[0,T]}2(T-s)^{1/2}s^{-1/2}ds \\
&= 4c^2|x-z|^2\int_{[0,T]}(T/s-1)^{1/2}ds \\
&= 8c^2T|x-z|^2 = d(T)|x-z|^2, \qquad (20.10)
\end{aligned}$$

say (Exercise 1). ■

Lemma 2 *Let f be a continuous function on $\mathbb{R}$ with compact support and $\{B_s\}$ a standard Brownian motion. Then,*

$$\int_{[0,t]}(\int_{(-\infty,\infty)} f(u)\mathbf{1}_{(u,\infty)}(B_s)du)dB_s = \int_{(-\infty,\infty)} f(u)(\int_{[0,t]}\mathbf{1}_{u,\infty)}(B_s)dB_s)du. \tag{20.11}$$

Proof Let the support of f be $[a, b]$. Define the partitions $\Pi_n : t_{j,n} = a + j(b - a)/2^n;\ j = 0, 1, \ldots, 2^n (n = 1, 2, \ldots)$, of $[a, b]$, $n = 1, 2, \ldots$. As shown in Theorem 20.1 and Lemma 1, $u \to \int_{[0,t]} \mathbf{1}_{u,\infty)}(B_s)dB_s$ is continuous almost surely, uniformly for all t in a bounded interval. Hence the right side of (20.11) is the limit of

$$\sum_{0\le j\le 2^{n-1}}(b-a)2^{-n}f(t_{j,n})(\int_{[0,t]}\mathbf{1}_{(t_{j,n},\infty)}(B_s)dB_s) = \int_{[0,t]}F_n(B_s)dB_s,$$

where

$$F_n(y) := \sum_{0\le j\le 2^{n-1}}(b-a)2^{-n}f(t_{j,n})\mathbf{1}_{(t_{j,n},\infty)}(y).$$

The sequence of Riemann sums F_n is bounded and converges uniformly to $F(y) = \int_{[0,t]} f(u)\mathbf{1}_{(u,\infty)}(y)du$. This implies that the sequence of stochastic integrals $\int_{[0,t]} F_n(B_s)dB_s$ converges almost surely and in L^1 to $\int_{[0,t]} F(B_s)dB_s$, namely, the left side of (20.11). Hence the two sides of (20.11) are equal almost surely. ■

In (20.4), replacing $\{B_s\}$ by $\{-B_s\}$, one obtains for the case $x = 0$, (20.4) changes to the corresponding expression for the local time at zero, $\tilde{\ell}(t, 0)$, say, given by

$$(-B_t)^+ + \int_{[0,t]} \mathbf{1}_{(-\infty,0]}(B_s)dB_s = \tilde{\ell}(t, 0). \tag{20.12}$$

Since the time during $[0, t]$ that $\{B_s : s \geq 0\}$ spends at zero is the same as that spent by $\{-B_s :\geq 0\}$,

$$\ell(t, 0) = \tilde{\ell}(t, 0) \quad \text{almost surely,}$$

so that by adding the two expressions (20.4) and (20.12), one arrives at

Corollary 20.2 (Tanaka's Formula) *Let $B = \{B_t : t \geq 0\}$ be a standard Brownian motion starting at zero. Then one has*

$$2\ell(t, 0) = |B_t| - \int_{[0,t]} sgn(B_s)dB_s \quad \textit{almost surely,}$$

or, equivalently, changing B to $-B$,

$$2\ell(t, 0) = |B_t| + \int_{[0,t]} sgn(B_s)dB_s \quad \textit{almost surely.} \tag{20.13}$$

Remark 20.2 Convexity of the functions $x \to |x|$, $x \to x^+$ plays an essential role in the Tanaka (1963) formula. Both the local time theory and Tanaka theorem for Brownian motion presented here have a natural extension to semimartingales[3] and differences of convex functions, respectively, e.g., see Revuz and Yor (1999), Chapter VI.

We conclude this chapter with another amazing representation by Lévy of the local time $\ell(t) = \ell(t, 0)$ at zero. For the proof of Lévy's result, we will use an important lemma due to Skorokhod.[4] It provides a unique representation of a continuous function f as the difference between a continuous non-negative g_z, say, starting at an arbitrarily given $z \geq 0$, and a nondecreasing continuous function k which only grows on the set of zeros of g_z.

[3] Also see Appuhamillage et al. (2014), Ramirez et al. (2016) for consequences of continuity of local time for physics.

[4] Skorokhod (1961, 1962).

Lemma 3 (Skorokhod's Equation) *Let*[5] $f(t)$ *be a continuous real-valued function on* $[0,\infty)$, $f(0) = 0$, *and let* $z \geq 0$ *be given. Then there exists a unique continuous function* $k(t)$ *on* $[0,\infty)$ *such that (i)* k *is nondecreasing, (ii)* $k(0) = 0$, *and (iii)* $k(\cdot)$ *grows only on the zeros of the function* $g(t) := z + k(t) + f(t)$, *i.e.,* k *is flat off* $\{t : g(t) = 0\}$. *The function* k *is given by*

$$k(t) = \max[0, \max\{-f(s) - z : 0 \leq s \leq t\}], t \geq 0. \tag{20.14}$$

Existence [6] We will check that $k(\cdot)$ in (20.12) satisfies the desired criteria (i)–(iii). Criteria (i) and (ii) are obviously met. For (iii), first note that if $t_1 < t_2$, then

$$k(t_2) = \max[k(t_1), \max\{-f(s) - z : t_1 \leq s \leq t_2].$$

Consider now, for any $\varepsilon > 0$, the open set $\{s : g(s) > \varepsilon\}$, which is a countable union of disjoint open intervals. If (t_1, t_2) is one of these intervals, then

$$k(t_2) = \max[k(t_1), \max\{-f(s) - z : t_1 \leq s \leq t_2].$$

But

$$k(t_i) + f(t_i) + z = g(t_i) = \varepsilon (i = 1, 2),$$

and

$$\max\{-f(s) - z : t_1 \leq s \leq t_2\} = \max\{-g(s) + k(s) : t_1 \leq s \leq t_2\} \leq -\varepsilon + k(t_2).$$

But this says $k(t_2) \leq \max[k(t_1), k(t_2) - \varepsilon]$, which cannot be true unless $k(t_2) = k(t_1)$. Hence $k(\cdot)$ is flat on $\{t : g(t) > \varepsilon\}$. This being true for every $\varepsilon > 0$, (iii) is verified.

(Uniqueness). To prove uniqueness, suppose $\tilde{g}$ and $\tilde{k}$ also satisfy the conditions above, but $g \neq \tilde{g}$. Suppose there is a point $T > 0$ such that $g(T) > \tilde{g}(T)$. Let

$$t_0 := \sup\{0 \leq s < T, g(s) = \tilde{g}(s)\}.$$

Note that one must have $g(s) > \tilde{g}(s)$ on $(t_0, T]$, implying $k(s) > \tilde{k}(s)$ on $(t_0, T]$. Since $\tilde{g}(s) \geq 0$, one must have $g(s) > 0$ on $(t_0, T]$. Hence $k(s)$ is flat on $(t_0, T]$, so that $k(t_0) = k(T)$ and $g(t_0) = g(T)$. Since

$$g(t) - \tilde{g}(t) = k(t) - \tilde{k}(t) \quad \forall\, t,$$

[5] Ikeda and Watanabe (1989), pp. 113–116; Karatzas and Shreve (1991), p. 233.

[6] Karatzas and Shreve (1991), pp. 210–212.

this leads to the contradiction

$$0 < g(T) - \tilde{g}(T) = k(T) - \tilde{k}(T) = k(t_0) - \tilde{k}(T) \le k(t_0) - \tilde{k}(t_0) = 0.$$

Hence $g = \tilde{g}$ on $[0, \infty)$ and, therefore, also $k(t) = \tilde{k}(t)$ for all t. ■

Remark 20.3 Skorokhod (1961–1962)[7] used his famous equation to construct reflecting diffusions. The method extends to multidimensional diffusions as well. We will use it to derive a theorem due to Lévy. A different proof, using Lévy's representation of processes with independent increments, specialized to passage times τ_a of a standard Brownian motion to $a \ge 0$, may be found in Itô and McKean (1965), pp. 42–48.

Theorem 20.3 (Lévy) *Let*[8] $X = \{X(t); t \ge 0\}$ *be a standard Brownian motion starting at zero,* $Y(t) = \max\{X(s) : 0 \le s \le t\}$*, and* $\ell(t)$ *the local time of* $\{X(t)\}$ *at zero. Then* $\{(|X(t)|, 2\ell(t)) : t \ge 0\}$ *has the same distribution as* $\{(Y(t) - X(t), Y(t)) : t \ge 0\}$*. In particular,* $Y(t) - X(t)$ *has the same distribution as* $|X(t)|$*, and* $2\ell(t)$ *has the same distribution as* $Y(t)$*.*

Proof We will check that Tanaka's formula (20.13) provides the Skorokhod representation needed, with $z = 0$, $f(s) = -W_s = -2\ell(s, 0) + |B_s|$, where $\{W_s : s \ge 0\}$ is the standard Brownian motion starting at 0, as given by the stochastic integral in (20.13). That is,

$$g(t) = |B_t| = k(t) + f(t),$$

with $k(t) = 2\ell(t)$. Note that $k(0) = 0$, $k(t)$ increases with t, and it grows only on the set of zeros of $g(t) = |B_t|$. By the uniqueness of the Skorokhod representation,

$$k(t) = \max\{W_s : 0 \le s \le t\} = M_t,$$

say, $g(t) = M_t - W_t$, and $(g(t), k(t)) \equiv (|B_t|, 2\ell(t))$ has the same distribution as $(M_t - W_t, M_t)$. Now apply this distributional result to the process $\{X(s) : s \ge 0\}$ for $W_s = X(s)$. ■

Exercises

1. Show that the constant c in the expression $d(T)$ (at the end of the proof of Lemma 1) may be taken to be $(2\pi)^{-1/2}$. Using this and the known expression of $c(2)$ (see Proposition 19.7 in Chapter 19 on stochastic integration with respect

[7] See Skorokhod (1961, 1962).

[8] See Lévy (1948).

to martingales), find a numerical upper bound for $c(T)$, apart from the constant multiple T.

2. (*Occupation Time Formula*) Show that (20.1) may be reformulated as $\int_0^t g(B_s)ds = 2\int_{\mathbb{R}} g(x)\ell(t,x)dx$ a.s., for bounded, measurable functions g. [*Hint*: Use simple function approximations to g.]
3. Show that, with probability one, $\ell(t,0) = \lim_{\varepsilon\downarrow 0} \frac{|\{0\le s\le t: B_s\in(-\varepsilon,\varepsilon)\}|}{4\varepsilon}$, where $|A|$ denotes the Lebesgue measure of a measurable set A, and $\{B_s : s \ge 0\}$ is a standard Brownian motion starting at zero.
4. Use Tanaka's formula to show that the inverse function of $t \to \ell(t,0)$ is an increasing stable process, i.e., a stable subordinator. [*Hint*: See Example 3 in Chapter 5.]
5. Show that $\mathbb{E}\ell(t,0) = \sqrt{\frac{t}{2\pi}}$. [*Hint*: By the reflection principle $P(Y(t) > y) = 2P(B_t > y)$, $y \ge 0$. This formula may vary in the literature in accordance with Remark 20.1 on the factor of '2'.]

Chapter 21
Construction of One-Dimensional Diffusions by Semigroups

This chapter contains a treatment of Feller's seminal contribution to a comprehensive theory of one-dimensional diffusions.

In a series of classic papers between 1952 and 1960, William Feller essentially showed that all regular Markov processes on an interval state space with continuous sample paths are generated (in the sense of semigroup theory) by generalized second-order differential operators[1] of the form $D_m D_s^+ = \frac{d}{dm}\frac{d^+}{ds}$, with appropriate boundary conditions; see Remark 21.1 below. Here, s is a strictly increasing continuous function on the interval, $D_s^+ f(x) \equiv \frac{d^+}{ds} f(x) = \lim_{y \to x^+} \frac{f(y)-f(x)}{s(y)-s(x)}$ denotes a right-sided derivative, and m is a strictly increasing right continuous function. A comprehensive account of this theory along with a wealth of additional facts appears in the book by Itô and McKean (1965). A somewhat simpler account than theirs may be found in the books of Mandl (1968) and Freedman (1971). Mandl (1968) gives Feller's construction of semigroups generated by $D_m D_s^+$, while Freedman's (1971) chapter on diffusions provides the probabilistic counterpart, demonstrating that all regular diffusions are necessarily of this form. This chapter contains another treatment of Feller's seminal contribution to the theory of diffusions. Here the term "regular diffusion" is in the sense of Freedman (1971). That is,

Definition 21.1 A regular diffusion is a Markov process having continuous sample paths, which has positive probabilities of moving to the left and to the right

[1] Especially see Feller (1957) in this regard.

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_21

instantaneously, starting from any point in the interior. "Instantaneously" means in any time interval, however small.

To be on familiar ground, first consider classical second-order differential operators

$$L = \frac{1}{2}a(x)\frac{d^2}{dx^2} + b(x)\frac{d}{dx}, \tag{21.1}$$

where $a(\cdot)$ and $b(\cdot)$ are Borel measurable functions on an interval (r_0, r_1), bounded on compact intervals; $a(\cdot)$ is positive on (r_0, r_1) and bounded away from zero on compact intervals. Assume that zero lies in the interior of the interval (r_0, r_1), the state space (This may be achieved by a simple translation).

Remark 21.1 Much of Feller's seminal work was motivated by an interest in mathematical biology, especially genetics.[2] These and a number of other of other interesting examples appear in selected examples and exercises at the end of this chapter.

Write

$$\begin{aligned} s(x) &= \textstyle\int_0^x \exp\{-\int_0^y \frac{2b(r)}{a(r)}dr\}dy, \\ m(x) &= \textstyle\int_0^x \frac{2}{a(y)} \exp\{\int_0^y \frac{2b(r)}{a(r)}dr\}dy. \end{aligned} \tag{21.2}$$

Then

$$\frac{ds(x)}{dx} = \exp\{-\int_0^x \frac{2b(r)}{a(r)}dr\}, \quad \frac{dm(x)}{dx} = \frac{2}{a(x)}\exp\{\int_0^x \frac{2b(r)}{a(r)}dr\} \tag{21.3}$$

and

$$\begin{aligned} D_m D_s^+ f(x) &= D_m \frac{df(x)}{ds(x)} = D_m \frac{df(x)}{dx} / \frac{d^+ s}{dx} \\ &= D_m[f'(x)\exp\{\int_0^x \frac{2b(r)}{a(r)}dr\}] \\ &= \frac{d}{dx}[f'(x)\exp\{\int_0^x \frac{2b(r)}{a(r)}dr\}] / \frac{dm(x)}{dx} \\ &= \frac{1}{2}a(x)f''(x) + b(x)f'(x) = Lf(x), \quad x \in (r_0, r_1), \end{aligned} \tag{21.4}$$

for all twice-differentiable functions f on I.

[2] See the survey article by Peskir (2015) for a penetrating overview.

Writing, $I = [r_0, r_1]$, where $-\infty \le r_0 < r_1 \le \infty$, the interval $[r_0, r_1]$ represents the two-point compactification of (r_0, r_1) obtained by identifying (r_0, r_1) with a *finite* interval (a, b) by a homeomorphism and then extending this homeomorphism to $[r_0, r_1]$ and $[a, b]$, r_0 being the image of a, and r_1 being the image of b. We then write $C(I)$ for the set of all real-valued continuous functions on this compact space I. One may, in case r_0 or r_1 is infinite, identify $C(I)$ with the set of all real-valued continuous functions on I having finite limits at both end points. $C(I)$ is a Banach space under "sup" norm, and this is the norm we will use in this chapter unless otherwise specified.

For the general case, we define the following.

Definition 21.2 The *scale function* $s(\cdot)$ is a continuous strictly increasing function on (r_0, r_1). The *speed function* $m(\cdot)$ is a right-continuous strictly increasing function on (r_0, r_1). One may view $m(\cdot)$ as the distribution function for the Lebesgue-Stieltjes measure, denoted dm or $m(dx)$, and referred to as the *speed measure*: $m(x) = \int_{(0,x]} m(dz)$, $x \in \mathbb{R}$ (see BCPT[3] p. 228).

It will be assumed throughout that $m(0) = 0$ and 0 is an interior point of I. This is simply achieved by translation since adding a constant to the speed function does not change the speed measure. In particular, D_s^+ denotes the right-hand derivative, i.e.,

$$D_s^+ f(x) = \lim_{y \to x^+} \frac{f(y) - f(x)}{s(y) - s(x)}, r_0 < x < r_1.$$

To simplify the calculations a little, we shall also take $s(x) = x$. This choice of scale function is referred to as *natural scale*. In this case,

$$D_s^+ f(x) \equiv D_x^+ f(x) = f'(x^+).$$

Natural scale can always be achieved by a change of variables transforming $x \to s(x)$. In this case the derivative D_s^+ will be denoted D_x^+. Denote by $\mathcal{D}_L$ the *domain of the operator* $L = D_m D_s^+$ from $C(I)$ to $C(I)$, i.e.,

$$\mathcal{D}_L = \{f \in C(I) : Lf \in C(I)\}. \tag{21.5}$$

To be clear, on the assumed natural scale $s(x) = x$, $f \in \mathcal{D}_L \subset C(I)$ requires that the one-sided derivative $D_x^+ f(x) \equiv f'(x^+)$ exist for all $x \in I$ and that $D_x^+ f$ induces a signed measure[4] absolutely continuous with respect to the measure $m(dx)$ and having Radon–Nikodym derivative[5] $D_m f'$. In addition, for $f \in \mathcal{D}_L$, it

[3] Throughout, BCPT refers to Bhattacharya and Waymire (2016), A Basic Course in Probability Theory.

[4] See BCPT, pp. 6, 228.

[5] See BCPT p. 250.

is required that

$$Lf(x) := D_m D_x^+ f(x),\ r_0 < x < r_1, \tag{21.6}$$

exists, as well as their limits $Lf(r_i) = D_m D_x^+ f(r_i), i = 0, 1$, and that Lf, so defined, belongs to $C(I)$.

Remark 21.2 The notations $m(dx)$ and $dm(x)$ will both be used to denote the speed measures.

Lemma 1 *$\mathcal{D}_L$ is dense in $C(I)$.*

Proof Let $f \in C(I)$. Fix $\varepsilon > 0$. The goal is to approximate f uniformly by a function $h \in \mathcal{D}_L$, where the domain is as defined above. Keep in mind that such an approximate must have a right-hand derivative, $D_x^+ h$, that is the distribution function of a Lebesgue-Stieltjes measure having a continuous Radon–Nikodym derivative (density), $D_m D_x^+ h$, with respect to the speed measure m. First let us note that without loss of generality, for the approximation, it may be assumed that f is continuously differentiable from the start. To see this, begin with a polygonal approximation by a step function f_1 of f such that f_1 is constant at each endpoint of I, and $||f_1 - f|| < \varepsilon$. Then, let ϕ_ε be an infinitely differentiable function on $\mathbb{R}$ vanishing outside $(-\frac{\delta_\varepsilon}{2}, \frac{\delta_\varepsilon}{2})$ where $|f_1(x) - f_1(y)| < \varepsilon$ if $|x - y| \le \delta_\varepsilon$. Consider the extension $\tilde{f}_1$ (if necessary) of f_1 to $(-\infty, \infty)$ by letting it have the value $f_1(r_0)$ on $(-\infty, r_0)$ and the value $f(r_1)$ on (r_1, ∞). Let $\tilde{f}_2 = \tilde{f}_1 * \phi_\varepsilon$. Let f_2 be the restriction of $\tilde{f}_2$ to $[r_0, r_1]$ (in case r_1 is finite). Then, $||f_2 - f_1|| < \varepsilon$. Thus $||f_2 - f|| < 2\varepsilon$. So, let us assume that $f \in C(I)$ is smooth from the start. Next, let us find a function f_3 that is absolutely continuous with respect to the speed measure m and that has a continuous Radon–Nikodym derivative $D_m f_3$ with respect to m to approximate the derivative f' of f in the sense that

$$\int_I |f'(y) - f_3(y)| dy < \varepsilon. \tag{21.7}$$

To construct such a function f_3, reconsider the polygonal approximation f_1 to f' by step functions, let $\delta > 0$ be arbitrarily small. Fix a finite interval, say $J = [j_0, j_1] \subset (r_0, r_1)$, on which f_1 is constant. Let $\varphi_i, i = 0, 1$, be a pair of nonnegative continuous functions, vanishing off of $[j_0 - \delta, j_0]$ and $j_1, j_1 + \delta]$, respectively, such that $\int_I \varphi_i(x) m(dx) = 1, i = 0, 1$. Then, the function $j(x) = \int_{r_0}^x (\varphi_0(y) - \varphi_1(y)) m(dy), x \in I$, is continuous and satisfies

$$j(x) = \begin{cases} 1 & \text{on } [j_0, j_1] \\ 0 & \text{on } (-\infty, j_0 - \delta) \cup (j_1 + \delta, \infty), \end{cases}$$

$0 \le j(x) \le 1$ on $[j_0 - \delta] \cup [j_1, j_1 + \delta]$, and one has

$$\int_I |\mathbf{1}_J(x) - j(x)|dx < 2\delta.$$

One may, for example, take $\varphi_i = \frac{\psi_i}{\int_{\mathbb{R}} \psi_i(z)m(dz)}$ $(i = 0, 1)$, where ψ_i is a continuous nonnegative function, $\int_{\mathbb{R}} \psi_i(z)m(dz) > 0$ $(i = 0, 1)$, ψ_0 vanishes outside $[j_0 - \delta, j_0]$, and ψ_1 vanishes outside $[j_1, j_1 + \delta]$. Applying this procedure to each step interval of f_1 and taking linear combinations, with suitably small δ, one obtains f_3, as desired. Now define

$$h(x) = f(0) + \int_0^x \int_{r_0}^y D_m f_3(z)m(dz)dy, \ x \in I.$$

Then,

$$h \in \mathcal{D}_L, \quad D_m D_x^+ h(x) = D_m f_3(x), \ x \in I,$$

and

$$\begin{aligned} ||f - h|| &= \sup_{x \in I} |\int_0^x (f'(y) - \int_{r_0}^y D_m f_3(z)m(dz))dy| \\ &= \sup_{x \in I} |\int_0^x (f'(y) - f_3(y))dy| \\ &\le \int_I |f'(y) - f_3(y)|dy < \varepsilon. \end{aligned}$$

■

It will be important later to distinguish between different types of possible boundary behaviors of $m(\cdot)$ and $s(\cdot)$ (or x, when in the natural scale).

Definition 21.3 The boundary point r_i $(i = 0 \text{ or } 1)$ is *inaccessible* if

$$|\int_0^{r_i} m(x)dx| = \infty, \tag{21.8}$$

and *accessible* otherwise.

Recall that (see (21.2)) $m(x) < 0$ for $x < 0$, and $\int_0^x = -\int_x^0$ if $x < 0$. (This is the convention in (21.2) and elsewhere).

Definition 21.4 An accessible boundary r_i is *regular* if

$$|\int_{[0,r_i)} xdm(x)| < \infty, \tag{21.9}$$

and is an *exit boundary* otherwise.

Remark 21.3 Note that for the integral in (21.8) to be infinity, r_i must be infinity ($i = 0, 1$). Hence r_i is inaccessible if it is plus or minus infinity. Thus, r_i is regular if r_i and $m(r_i)$ are both finite, and r_i is an exit boundary if it is finite, but $m(r_i)$ is plus or minus infinity.

Definition 21.5 An inaccessible boundary r_i is *natural* if the integral in (21.9) is infinite (i.e., if and only if both integrals are infinite), and an *entrance bondary* if the integral in (21.9) is finite, but (21.8) is infinite.

Our first goal is to determine explicit conditions on s and m in terms of r_i to uniquely determine the existence of a strongly continuous contraction semigroup with generator $(L, \mathcal{D}_L)$.

The approach to achieve this is to show (by the Hille–Yosida Theorem) that

$$\lambda \in \rho(L) \text{ and } ||R_\lambda|| \le \frac{1}{\lambda}\ \forall \lambda > 0 \text{ iff } r_0, r_1 \text{ are inaccessible boundaries.}$$

Toward this end, fix $\lambda > 0$ and try to obtain a solution $u \in \mathcal{D}_L$ of

$$(\lambda - L)u = f \tag{21.10}$$

for an arbitrary $f \in C(I)$. This is accomplished by obtaining a *Green's function* $G(x, y)$ (also called a *fundamental solution*) for the operator $\lambda - L$ such that

$$u(x) = (Gf)(x) \equiv \int_{(r_0,r_1)} G(x, y) f(y) dm(y) \tag{21.11}$$

solves (21.10) for every $f \in C(I)$. This is carried out by completing the following steps:

(i) Construct, by successive approximation, a function $g(x) = g(x; \lambda)$ that satisfies

$$(\lambda - L)g(x) = 0 \quad r_0 < x < r_1, \tag{21.12}$$

is positive and decreasing on $(r_0, 0)$ and increasing on $(0, r_1)$.

(ii) Set

$$g_+(x) = g(x) \int_x^{r_1} g(y)^{-2} dy, \quad g_-(x) = g(x) \int_{r_0}^{x} g(y)^{-2} dy. \tag{21.13}$$

One notes (see Lemma 5 below) that g_+ is decreasing, g_- increasing, and $g_\pm$ solve (21.12).

(iii) Use g_+, g_-, to construct a Green's function, $G \propto g_- g_+$, in the usual manner (developed in detail below, see (21.44)).

These steps will be carried out in a sequence of lemmas that provide a proof of the following theorem; specifically, steps (i)–(iii) are proven in Lemmas 4, 5, and 7. In the process of proving the theorem, a number of results applicable to the accessible cases will be derived as well.

Theorem 21.1

(a) The operator L with domain $\mathcal{D}_L$ is the infinitesimal generator of a (strongly continuous) contraction semigroup on $C(I)$ if and only if r_0 and r_1 are inaccessible boundaries.

(b) This semigroup is Markovian and has a (Feller continuous) transition probability $p(t; x, dy)$ on I satisfying

$$\lim_{t\downarrow 0} \frac{1}{t}\{\sup_{x\in I} p(t; x, B_\varepsilon^c(x))\} = 0; \textit{ for all } \varepsilon > 0. \tag{21.14}$$

Here $B_\varepsilon(x) = \{y : |y - x| < \varepsilon\}$. As noted above, the proof is accomplished by a series of lemmas.

Definition 21.6 The *Wronskian* $W(v_1, v_2)$ of functions v_1, v_2 is defined by the following determinant:

$$W(v_1, v_2)(x) = (\frac{d}{dx}v_1(x))v_2(x) - v_1(x)\frac{d}{dx}v_2(x) = \det\begin{bmatrix} \frac{d}{dx}v_1(x) & \frac{d}{dx}v_2(x) \\ v_1(x) & v_2(x) \end{bmatrix} \tag{21.15}$$

Lemma 2 *Suppose that v_1, v_2 are both solutions of $(\lambda - L)v = 0$. Then the Wronskian (21.15) is independent of x. It is identically zero if and only if v_1, v_2 are linearly dependent.*

Proof The second part of the lemma is clear from the definition of W. For the first part, we will show that the variation of W is zero on $(x, y] \subset I$. Note that $W(x) \equiv v_1'(x)v_2(x) - v_1(x)v_2'(x)$ is of bounded variation on finite subintervals of (r_0, r_1). Hence, for $r_0 < y < x < r_1$,

$$\begin{aligned} W(x) - W(y) &= \int_{(y,x]} dW(z) \\ &= \int_{(y,x]} d(v_1'(z)v_2(z)) - \int_{(y,x]} d(v_2'(z)v_1(z)) \\ &= \int_{(y,x]} (v_1'v_2'(z))dz + \int_{(y,x]} v_2(z)\frac{dv_1'(z)}{dm(z)}dm(z) \\ &\quad - \int_{(y,x]} (v_1'v_2'(z))dz - \int_{(y,x]} v_1(z)\frac{dv_2'(z)}{dm(z)}dm(z) \end{aligned}$$

$$
\begin{aligned}
&= \int_{(y,x]} v_2(z)(Lv_1(z))dm(z) - \int_{(y,x]} v_1(z)(Lv_2(z))dm(z) \\
&= \int_{(y,x]} v_2(z)\lambda v_1(z)dm(z) - \int_{(y,x]} v_1(z)\lambda v_2(z)dm(z) \\
&= 0.
\end{aligned}
$$

■

Lemma 3 *Let v_1, v_2 be two solutions of $(\lambda - L)v = 0$, such that $W(v_1, v_2) \neq 0$. Then every solution v of this equation is a unique linear combination of v_1 and v_2. In particular, v_1, v_2 are independent solutions if and only if $W(v_1, v_2) \neq 0$.*

Proof The second part of the lemma is contained in Lemma 2. For the first part, let v_1, v_2 be two linearly independent solutions and v an arbitrary solution of

$$(\lambda - L)v = 0.$$

There exist unique constants c_1, c_2 such that

$$c_1 v_1(0) + c_2 v_2(0) = v(0),$$

$$c_1(\frac{d}{dx}v_1)(0) + c_2(\frac{d}{dx}v_2)(0) = (\frac{d}{dx}v)(0),$$

since $\left(v_1(0), \left(\frac{d}{dx}v_1\right)(0)\right)$, and $\left(v_2(0), \left(\frac{d}{dx}v_2\right)(0)\right)$ are linearly independent in $\mathbb{R}^2$ by Lemma 2. We claim $u_0 := c_1 v_1 + c_2 v_2 = v$. To prove this note that $(L - \lambda)u_0 = 0$, $u_0(0) = v(0)$, and $u_0'(0) = v'(0)$. Therefore, $W(u_0, v)(0) = 0$, and hence, $W(u_0, v)(z) = 0$ for all z. This implies, by Lemma 2, that $v = cu_0$ for some constant c. But then $c = 1$ since this is the only constant for which $v(0) = cu_0(0)$. ■

Lemma 4 *There is a function $g \in \mathcal{D}_L$ such that*

$$Lg(x) = \frac{d}{dm}\frac{d}{dx}g(x) = \lambda g(x), \tag{21.16}$$

and $1 + \lambda \int_0^x m(y)dy \leq g(x) \leq \exp(\lambda \int_0^x m(y)dy)$.

Proof Throughout we use the convention for $x < 0$,

$$\int_{(0,x]} h(z)dm(z) = -\int_{[x,0)} h(z)dm(z), \quad \int_0^x h(z)m(dz) = -\int_x^0 h(z)m(dz). \tag{21.17}$$

Recursively define the functions

$$g^0(x) \equiv 1, \quad g^{n+1}(x) = \int_0^x (\int_{(0,y]} g^n(z)dm(z))dy, \ (n = 1, 2, \dots). \tag{21.18}$$

Then one has

$$Lg^{n+1} = g^n(x) \quad (n = 0, 1, 2 \dots). \tag{21.19}$$

By induction check that, for each $n \geq 1$, $g^n(x)$ is positive for $x \neq 0$ ($g^n(0) = 0$), strictly decreasing for $x < 0$, and strictly increasing for $x > 0$. Write

$$g(x) = g(x; \lambda) = \sum_{n=0}^{\infty} \lambda^n g^n(x). \tag{21.20}$$

At least formally, $Lg(x) = \lambda g(x)$ in view of (21.19). To prove the desired convergence properties, we use induction on n go get, for $x > 0$,

$$\begin{aligned} g^{n+1}(x) &= \int_0^x (\int_{(0,y]} g^n(z)dm(z))dy \\ &\leq \int_0^x g^n(y)m(y)dy = \int_0^x g^n(y_n)m(y_n)dy_n \\ &\leq \int_0^x (\int_0^{y_n} g^{n-1}(y_{n-1})m(y_{n-1})dy_{n-1})m(y_n)dy_n \\ &\vdots \\ &\leq \int_0^x \int_0^{y_n} \cdots \int_0^{y_1} m(y_0)m(y_1)\cdots m(y_n)dy_0dy_1\cdots dy_n \\ &= \frac{(g^1(x))^{n+1}}{(n+1)!}, \end{aligned} \tag{21.21}$$

where $g^1(z) = \int_0^z m(y)dy$, i.e., $dg^1(z) = m(z)dz$.

Similarly, for $x < 0$,

$$\begin{aligned} g^{n+1}(x) &= \int_x^0 (\int_{[y,0)} g^n(x)dm(z))dy \\ &\leq -\int_x^0 g^n(y)m(y)dy \\ &\vdots \\ &\leq \frac{g^1(x))^{n+1}}{(n+1)!}, \end{aligned}$$

where $g^1(z) = -\int_z^0 m(y)dy > 0$, i.e., $dg^1 = m(z)dz$. As a consequence

$$\left|\frac{dg^{n+1}(x)}{dx}\right| = \left|\int_{(0,x]} g^n(y)dm(y)\right| \le \left|\int_{(0,x]} \frac{(g^1(y))^n}{n!}dm(y)\right|$$
$$\le \frac{(g(x))^n}{n!}|m(x)|, \quad (n = 1, 2, \dots). \tag{21.22}$$

It follows from (21.21) that the function g in (21.20) converges absolutely, uniformly on compact subsets of (r_0, r_1), and that

$$g(x) \le \exp\{\lambda g^1(x)\} \tag{21.23}$$

It follows from (21.22) that the series of derivatives $\sum_0^\infty \lambda^n \frac{d}{dx} g^n(x)$ converges absolutely, uniformly on compacts, and

$$\sum_{n=0}^{\infty} \left|\lambda^n \frac{dg^n(x)}{dx}\right| = \sum_{n=1}^{\infty} \left|\lambda^n \frac{dg^n(x)}{dx}\right| \le |\lambda m(x)| e^{\lambda g^1(x)}. \tag{21.24}$$

Hence g is continuously differentiable and one has

$$\frac{d}{dx} g(x) = \sum_{n=1}^{\infty} \lambda^n \int_{(0,x]} g^{n-1}(z)dm(z) = \lambda \int_{(0,x]} g(z)dm(z). \tag{21.25}$$

Hence,

$$Lg(x) = \frac{d}{dm}\frac{d}{dx} g(x) = \lambda g(x), \tag{21.26}$$

and by (21.21) (and the obvious fact that $g \ge g^0 + \lambda g^1$),

$$1 + \lambda g^1(x) \le g(x) \le \exp\{\lambda g^1(x)\}. \tag{21.27}$$

■

Example 1 (1/2 Laplacian) In the case $s(x) = x, m(x) = 2x$ on $(-\infty, \infty)$, $L = \frac{1}{2}\frac{d^2}{dx^2}$, and one obtains from (21.20) that

$$g(x; \lambda) = \sum_{n=0}^{\infty} \frac{2^n \lambda^n x^{2n}}{(2n)!} = \cosh(\sqrt{2\lambda}x),$$

so that

$$g_+(x) = \cosh(\sqrt{2\lambda}x)(\frac{1}{\sqrt{2\lambda}} - \frac{\tanh(\sqrt{2\lambda}x)}{\sqrt{2\lambda}}) = \frac{e^{-\sqrt{2\lambda}x}}{\sqrt{2\lambda}},$$

and

$$g_-(x) = \cosh(\sqrt{2\lambda}x)(\frac{1}{\sqrt{2\lambda}} + \frac{\tanh(\sqrt{2\lambda}x)}{\sqrt{2\lambda}}) = \frac{e^{\sqrt{2\lambda}x}}{\sqrt{2\lambda}},$$

and, therefore, $W(g_-, g_+) = \frac{2}{\sqrt{2\lambda}}$. Thus, in accordance with (21.44) below,

$$G(x, y) = \frac{1}{2\sqrt{2\lambda}} e^{-\sqrt{2\lambda}|y-x|}, \ -\infty < x, y < \infty.$$

Exercise 14 contains another illustrative calculation along these lines.

Lemma 5 *The functions g_+, g_- defined in (21.13) are, respectively, strictly decreasing and strictly increasing on I, and each solves (21.26).*

Proof In view of the first inequality in (21.27), one has

$$\begin{aligned} g_+(x) &\leq g(x) \int_x^{r_1} (1 + \lambda g^1(z))^{-2} dz \\ &= g(x) \int_x^{r_1} (1 + \lambda g^1(z))^{-2} \frac{dg^1(z)}{m(z)} \\ &\leq \frac{g(x)}{|m(x)|} \int_x^{r_1} (1 + \lambda g^1(z))^{-2} dg^1(z) \qquad (21.28) \\ &= \frac{g(x)}{\lambda |m(x)|} \{(1 + \lambda g^1(x))^{-1} - (1 + \lambda g^1(r_1))^{-1}\} \\ &\leq \frac{g(x)}{\lambda |m(x)|} (1 + \lambda g^1(x))^{-1}, \end{aligned}$$

which is finite (for $x > 0$ and, therefore, for all x). Similarly,

$$\begin{aligned} g_-(x) &\leq \frac{g(x)}{\lambda |m(x)|} \{(1 + \lambda g^1(x))^{-1} - (1 + \lambda g^1(r_0))^{-1}\} \\ &\leq \frac{g(x)}{\lambda |m(x)|} (1 + \lambda g^1(x))^{-1} \qquad (21.29) \end{aligned}$$

is finite. Now let us check that g_+ and g_- also satisfy the eigenvalue problem of the form (21.26), or, equivalently, (21.12). Using the definition of g_+ and (21.26) for g, one has

$$
\begin{aligned}
&\lambda \int_0^x g_+(y)dm(y) \\
&= \lambda \int_0^x g(y)(\int_y^{r_1} \frac{1}{g^2(z)} dz)dm(y) \\
&= \int_0^x \frac{d}{dm} g'(y^+)(\int_y^{r_1} \frac{1}{g^2(z)} dz)dm(y) \\
&= \int_0^x (\int_y^{r_1} \frac{1}{g^2(z)} dz)dg'(y) \\
&= g'(x^+) \int_x^{r_1} \frac{1}{g^2(z)} dz - \int_0^x g'(y^+) \frac{d}{dy}(\int_y^{r_1} \frac{1}{g^2(z)} dz)dy \\
&= g'(x^+) \int_x^{r_1} \frac{1}{g^2(z)} dz - \frac{1}{g(x)} + \frac{1}{g(0)}. \qquad (21.30)
\end{aligned}
$$

Now, since $g'(0^+) = 0$, one has

$$
\begin{aligned}
g'_+(0^+) &= \lim_{x \to 0^+} \frac{1}{x}(g(x) \int_x^{r_1} \frac{dz}{g^2(z)} - g(0) \int_0^{r_1} \frac{dz}{g^2(z)}) \\
&= - \lim_{x \to 0^+} \frac{g(x) \int_0^x \frac{dz}{g^2(z)}}{x} = -\frac{1}{g(0)}. \qquad (21.31)
\end{aligned}
$$

Thus, using this and the definition of g_+, the last line of (21.30) is precisely $g'_+(x^+) - g'_+(0^+)$. Thus,

$$
g'_+(x^+) - g'_+(0^+) = \lambda \int_0^x g_+(y)dm(y). \qquad (21.32)
$$

Hence, one has that $Lg_+ = \lambda g_+$. Similarly, one has that

$$
Lg_-(x) = \frac{d}{dm}\frac{d}{dx} g_-(x) = \lambda g_-(x). \qquad (21.33)
$$

Next note that g is strictly positive, and $\frac{d}{dx} g(x)$ is increasing in x (see (21.25), (21.27)) so that

$$
\begin{aligned}
\frac{d}{dx} g_+(x) &= (\frac{d}{dx} g(x)) \int_x^{r_1} g(z)^{-2} dz - g(x)^{-1} \\
&< \int_x^{r_1} g(z)^{-2} \frac{d}{dz} g(z) dz - g(x)^{-1} = -g(r_1)^{-1} \le 0. \qquad (21.34)
\end{aligned}
$$

Here $g(r_1) = \lim_{x \to r_1} g(x)$. Hence, g_+ is *decreasing (strictly)*. Similarly,

$$\frac{d}{dx}g_-(x) > g(r_0)^{-1} \geq 0, \quad (g(r_0) = \lim_{x\to r_0} g(x)) \tag{21.35}$$

so that g_- is *strictly increasing*. ■

Next let v_1, v_2 be two solutions of (21.12), i.e.,

$$(\lambda - L)v_i(x) = 0, \qquad \text{for } r_0 < x < r_1 \quad (i = 1, 2). \tag{21.36}$$

The next order of business is to study the boundary behavior of $g_\pm$, $\frac{dg_\pm}{dx}$. The following table of values at r_i $(i = 1, 2)$ are obtained as limits as $x \to r_i$ with $x \in (r_0, r_i)$.

Lemma 6 *One has the following boundary behavior for $g_\pm$ given in Table 21.1.*

Proof

(I). Since $\frac{d}{dm}\frac{d}{dx}g_+ = \lambda g_+$, one has for $0 < x < x_0 < r_1$, integrating with respect to $m(dz)$,

$$(\frac{d}{dx}g_+)(x_0) - (\frac{d}{dx}g_+)(x) = \lambda \int_{(x,x_0]} g_+(z)dm(z).$$

Next, integrating with respect to dx,

$$(x_0 - x)g'_+(x_0) - g_+(x_0) + g_+(x) = \lambda \int_{(x,x_0]} (\int_{(z,x_0]} g_+(y)m(dy))dz. \tag{21.37}$$

Therefore,

$$\begin{aligned} g(x) &= g_+(x_0) - (x_0 - x)g'_+(x_0) + \lambda \int_{(x,x_0]} (\int_{(z,x_0]} g_+(y)m(dy))dz \\ &\geq \lambda \int_{(x,x_0]} (\int_{(z,x_0]} g_+(y)m(dy))dz. \end{aligned}$$

Letting $x_0 \uparrow r_1$, one gets (since $\frac{d}{dx}g_+ < 0$)

$$g_+(x) \geq \lambda \int_{(x,r_1]} (\int_{(z,r_1)} g_+(y)dm(y))dz \geq \lambda g_+(r_1) \int_{(z,r_1]} (m(r_1) - m(z))dz. \tag{21.38}$$

If r_1 is *exit* or *natural*, then

$$\infty = \int_{(0,r_1)} y dm(y) = \int_0^{r_1} (m(r_1) - m(y))dy$$

so that, by (21.38), $g_+(r_1)$ must vanish. If r_1 is *regular*, then r_1, $m(r_1)$ are both finite, g is bounded away from zero and infinity on $[x, r_1)$ and, therefore,

$$g_+(x) \equiv g(x) \int_{(x,r_1]} g^{-2}(y)dy \to 0$$

Table 21.1 Boundary Behavior of $g_\pm$

		Natural	Exit	Entrance	Regular
I	$g_+(r_1)$	0	0	> 0	0
II	$g_-(r_1)$	∞	$< \infty$	∞	$< \infty$
III	$\left(\frac{d}{dx}g_+\right)(r_1)$	0	< 0	0	≤ 0
IV	$\left(\frac{d}{dx}g_-\right)(r_1)$	∞	∞	$< \infty$	$< \infty$
V	$g_+(r_0)$	∞	$< \infty$	∞	$< \infty$
VI	$g_-(r_0)$	0	0	> 0	0
VII	$\left(\frac{d}{dx}g_+\right)(r_0)$	$-\infty$	$-\infty$	$> \infty$	$> -\infty$
VIII	$\left(\frac{d}{dx}g_-\right)(r_0)$	0	> 0	0	≥ 0

as $x \uparrow r_1$. This validates the first row in Table 21.1, except for the case of an *entrance boundary*, which we establish as follows. The function $g_+(x)$ is decreasing, so that $\frac{d}{dx}g_+ \leq 0$; also, for $x > 0$,

$$(\frac{d}{dx}g_\pm)(x) = (\frac{d}{dx}g_\pm)(0) + \lambda \int_{(0,x]} g_\pm(z)dm(z), \tag{21.39}$$

which shows that $\frac{d}{dx}g_\pm$ is increasing, and, in particular, the limit at r_1, $(\frac{d}{dx}g_+)(r_1)$ exists and is ≤ 0. If r_1 is inaccessible, then

$$g(r_1) \geq (1 + \lambda g^1(r_1)) = \infty,$$

so that $g(r_1)^{-1} = 0$. Letting $x \uparrow r_1$, in the first equality in (21.30), one gets $(\frac{d}{dx}g_+)(r_1) \geq 0$ (since $(\frac{d}{dx}g)(x) \geq 0$ for $x > 0$). Hence $(\frac{d}{dx}g_+)(r_1) = 0$ if r_1 is inaccessible. Now, using this, one has

$$\begin{aligned}
&-\log g_+(r_1) + \log g_+(0) \\
&= \int_0^{r_1} -\frac{(\frac{d}{dx}g_+)(x)}{g_+(x)}dx \\
&= \int_0^{r_1} \frac{(\frac{d}{dx}g_+)(r_1) - (\frac{d}{dx}g_+)(x)}{g_+(x)}dx, \qquad ((\frac{d}{dx}g_+)(r_1) = 0) \\
&= \int_0^{r_1} \frac{\lambda}{g_+(x)}(\int_{(x,r_1)} g_+(y)dm(y))dx, \qquad (\frac{d}{dm}\frac{d}{dm}g_+) = \lambda g_+ \\
&\leq \lambda \int_0^{r_1} \frac{g_+(x)}{g_+(x)}(\int_{(x,r_1)} dm(y))dx = \lambda \int_0^{r_1} (\int_{(x,r_1)} dm(y))dx \\
&= \lambda \int_{(0,r_1)} (\int_0^y dx)dm(y) = \lambda \int_{(0,r_1)} ydm(y) < \infty,
\end{aligned} \tag{21.40}$$

since r_1 is an entrance boundary. This shows $g_+(r_1) \neq 0$.

(II). By the Lemma 2, there exist unique real numbers c_1, c_2 such that $g(x) = c_1 g_+(x) + c_2 g_-(x)$. Clearly, c_1, c_2 are nonzero (since $g(x)$ is strictly decreasing for $x < 0$ and strictly increasing for $x > 0$) and cannot both be negative (since $g(x) > 0$ for $x \neq 0$). Indeed, both are positive. For suppose $c_1 < 0$ and $c_2 > 0$; then $\frac{d}{dx} g(x) = c_1 \frac{d}{dx} g_+(x) + c_2 \frac{d}{dx} g_-(x) > 0$ for all x, but this is not true. Similarly, $c_1 > 0$ and $c_2 < 0$ would imply that $\frac{d}{dx} g(x)$ is negative for all x. Next note that $g(r_1) = \infty$ if and only if $g^1(r_1) = \infty$ if and only if $g_-(r_1) = \infty$. Hence, $g_-(r_1) = \infty$ if and only if r_1 is inaccessible.

(III). In establishing (I), we have already proved that $(\frac{d}{dx} g_+)(r_1) = 0$ if r_1 is inaccessible (i.e., *natural* or *entrance*). The assertion $(\frac{d}{dx} g_+)(r_1) \leq 0$ if r_1 is *regular*, is obvious. Assume now that r_1 is *exit*. By Lemma 1,

$$\begin{aligned} 0 &< (\frac{d}{dx} g_-)(0) g_+(0) - g_-(0)(\frac{d}{dx} g_+)(0) = W(g_-, g_+) \\ &= \lim_{x \uparrow r_1} \{(\frac{d}{dx} g_-)(x) g_+(x) - g_-(x)(\frac{d}{dx} g_+)(x)\}. \end{aligned} \tag{21.41}$$

Now, since $g_+(r_1) = 0$ (by (I)),

$$\begin{aligned} (\frac{d}{dx} g_-)(x) g_+(x) &= -(\frac{d}{dx} g_-)(x) \int_x^{r_1} \frac{dg_+}{dz}(z) dz, \\ &= -\int_x^{r_1} (\frac{d}{dx} g_-(x))(\frac{d}{dz} g_+(z)) dz \\ &\leq -(\frac{d}{dx} g_+(x)) \int_x^{r_1} \frac{d}{dz} g_-(z) dz, \end{aligned} \tag{21.42}$$

since $\frac{d}{dx} g_\pm(x) \uparrow$ as $x \uparrow$ (see (21.39)). Recalling that $g_-(r_1) < \infty$ (by (II)), and $\frac{d}{dx} g_+(x)$ goes to a finite limit as $x \uparrow r_1$ (since $\frac{d}{dx} g_+(x)$ is negative and increasing), one gets from (21.42)

$$\begin{aligned} (\frac{d}{dx} g_-(x)) g_+(x) &\leq (-\frac{d}{dx} g_+(x))(g_-(r_1) - g_-(x)) \\ &\to 0 \text{ as } x \uparrow r_1. \end{aligned}$$

Using this in (21.41), we have

$$-g_-(r_1)(\frac{d}{dx} g_+)(r_1) > 0,$$

which implies that $(\frac{d}{dx} g_+(r_1))(r_1) < 0$.

(IV). One has $g_-(y) = g_-(0) + \int_0^y \frac{d}{dz} g_-(z) dz \geq \int_0^y \frac{d}{dz} g_-(z) dz \geq \int_0^y (\frac{d}{dz} g_-)(0) dz = y(\frac{d}{dz} g_-)(0)$. Hence

$$
\begin{aligned}
(\frac{d}{dx}g_-)(x) - (\frac{d}{dx}g_-)(0) &= \lambda \int_{(0,x]} g_-(y)dm(y) \\
&\geq \lambda \int_{(0,x]} y(\frac{d}{dz}g_-)(0)dm(y) \\
&= \lambda(\frac{d}{dz}g_-)(0) \int_{(0,x]} ydm(y).
\end{aligned}
\tag{21.43}
$$

Letting $x \uparrow r_1$, one gets (since $(\frac{d}{dz}g_-)(0) > 0$),

$$
(\frac{d}{dx}g_-)(r_1) = \infty,
$$

if r_1 is *natural* or *exit*. If r_1 is *regular*, then $r_1 < \infty$, $m(r_1) < \infty$, $g^1(r_1) < \infty$, and, therefore, $g(r_1) < \infty$ and $(\frac{d}{dx}g)(r_1) < \infty$. Since $(\frac{d}{dx}g)(r_1) = c_1(\frac{d}{dx}g_+)(r_1) + c_2(\frac{d}{dx}g_-)(r_1)$ for positive constants c_1, c_2, it follows that $(\frac{d}{dx}g_-)(r_1) < \infty$. If r_1 is *entrance*, then $g_+(r_1) > 0$ (by (I)), and if $(\frac{d}{dx}g_-)(r_1) = \infty$, then the right side of (21.41) would be $+\infty$, a contradiction since $W(g_-, g_+) = W(g_-, g_+)(0)$ is finite.

The proof of (I)–(IV) is complete; the proof of the boundary behavior at r_0 is entirely analogous. ■

Let us now show that the desired *Green's function* for $\lambda - L$ is given as follows:

Lemma 7 *The Green's function for $\lambda - L$ is given by*

$$
G(x, y) = \begin{cases} \frac{g_-(x)g_+(y)}{W(g_-,g_+)} & \text{if } r_0 < x < y < r_1 \\ G(y, x) & \text{if } r_0 < y \leq x < r_1. \end{cases}
\tag{21.44}
$$

For every bounded measurable real-valued function f on I,

$$
(Gf)(x) := \int_{(r_0,r_1)} G(x, y)f(y)dm(y) \quad r_0 < x < r_1.
\tag{21.45}
$$

The integral is bounded and for all $f \in C(I)$,

$$
(\lambda - L)Gf(x) = f(x) \quad r_0 < x < r_1.
\tag{21.46}
$$

Proof To show that the integral on the right converges, write $W = W(g_-, g_+)$ and note that

$$
\begin{aligned}
&W(Gf)(x)\\
&= g_+(x)\int_{(r_0,x]} g_-(y)f(y)dm(y) + g_-(x)\int_{(x,r_1)} g_+(y)f(y)dm(y)\\
&= \lambda^{-1}g_+(x)\int_{(r_0,x]}(\frac{d}{dm}\frac{d}{dy}g_-(y))f(y)dm(y)\\
&\quad +\lambda^{-1}g_-(x)\int_{(x,r_1)}(\frac{d}{dm}\frac{d}{dy}g_+(y))f(y)dm(y),\\
&= \lambda^{-1}g_+(x)\int_{(r_0,x]} f(y)d(\frac{d}{dy}g_-(y)) + \lambda^{-1}g_-(x)\int_{(x,r_1)} f(y)d(\frac{d}{dy}g_+(y)),
\end{aligned}
\tag{21.47}
$$

which is bounded in magnitude by

$$
\begin{aligned}
&\lambda^{-1}g_+(x)|\{(\frac{d}{dy}g_-(y))_{y=x} - (\frac{d}{dy}g_-(y))_{y=r_0}\}|||f||_\infty\\
&\quad +\lambda^{-1}g_-(x)|\{(\frac{d}{dy}g_+(y))_{y=r_1} - (\frac{d}{dy}g_+(y))_{y=x}\}|||f||_\infty.
\end{aligned}
\tag{21.48}
$$

To prove, (21.46), from (21.47) one gets

$$
\begin{aligned}
\frac{d}{dx}W(Gf)(x) &= (\frac{d}{dx}g_+)(x)\int_{(r_0,x]} g_-(y)f(y)dm(y)\\
&\quad +(\frac{d}{dx}g_-)(x)\int_{(x,r_1)} g_+(y)f(y)dm(y),
\end{aligned}
$$

so that

$$
\begin{aligned}
&\frac{d}{dm}\frac{d}{dx}W(Gf)(x)\\
&= (\frac{d}{dm}\frac{d}{dx}g_+)(x)\int_{(r_0,x]} g_-(y)f(y)dm(y)\\
&\quad +(\frac{d}{dx}g_+)(x)g_-(x)f(x) + (\frac{d}{dm}\frac{d}{dx}g_-)(x)\int_{(x,r_1)} g_+(y)f(y)dm(y)\\
&\quad -(\frac{d}{dx}g_-)(x)g_+(x)f(x)\\
&= \lambda g_+(x)\int_{(r_0,x]} g_-(y)f(y)dm(y) + \lambda g_-(x)\int_{(x,r_1)} g_+(y)f(y)dm(y) - Wf(x)\\
&= \lambda W(Gf)(x) - Wf(x).
\end{aligned}
$$

■

The next lemma deals with the boundary behavior of Gf.

Lemma 8 *Let $f \in C(I)$.*

(a) Then $Gf \in C(I)$.
(b) If r_i is accessible, then $(Gf)(r_i) := \lim_{x\to r_i}(Gf)(x) = 0$
(c) If r_i is natural, then $(Gf)(r_i) = \lambda^{-1}f(r_i)$ $(i = 1, 2)$.

Proof Take $r_i = r_1$. Consider first the case: *f vanishes in a neighborhood of r_1*. Let $f(x) = 0$ for $x > r$ $(r < r_1)$. Then

$$W(Gf)(x) = g_+(x)\int_{(r_0,x]} g_-(y)f(y)dm(y) + g_-(x)\int_{(x,r_1)} g_+(y)f(y)dm(y) \tag{21.49}$$

$$\to g_+(r_1)\int_{(r_0,r]} g_-(y)f(y)dm(y) \text{ as } x \uparrow r_1.$$

If r_1 is accessible (i.e., *regular* or *exit*) or *natural*, then $g_+(r_1) = 0$ and $(Gf)(r_1) = 0$. Next consider the case: $f \equiv 1$. In this case (see the last equality in (21.47)):

$$\begin{aligned}
&W(Gf)(x)\\
&= \lambda^{-1}g_+(x)((\frac{d}{dx}g_-)(x) - (\frac{d}{dx}g_-)(r_0))\\
&+\lambda^{-1}g_-(x)((\frac{d}{dx}g_+)(r_1) - (\frac{d}{dx}g_+)(x))\\
&= \lambda^{-1}W - \lambda^{-1}g_+(x)(\frac{d}{dx}g_-)(r_0) + \lambda^{-1}g_-(x)(\frac{d}{dx}g_+)(r_1).
\end{aligned} \tag{21.50}$$

In case r_1 is *accessible*, then $g_+(r_1) = 0$, $g_-(r_1) < \infty$, and one has

$$\begin{aligned}
&W\lim_{x\uparrow r_1}(Gf)(x)\\
&= \lambda^{-1}W + \lambda^{-1}g_-(r_1)(\frac{d}{dx}g_+)(r_1)\\
&= \lambda^{-1}\lim_{x\uparrow r_1}((\frac{d}{dx}g_-)(x)g_+(x) - g_-(x)(\frac{d}{dx}g_+)(x) + g_-(x)(\frac{d}{dx}g_+)(x))\\
&= \lambda^{-1}\lim_{x\uparrow r_1}(\frac{d}{dx}g_-)(x)g_+(x).
\end{aligned} \tag{21.51}$$

Now one has

$$\begin{aligned}
(\frac{d}{dx}g_-)(x) &= (\frac{d}{dx}g_-)(0) + \int_{(0,x]}\lambda g_-(y)dm(y)\\
&\le (\frac{d}{dx}g_-)(0) + \lambda g_-(r_1)m(x),
\end{aligned}$$

so that

$$\begin{aligned}
\lim_{x\uparrow r_1}(\frac{d}{dx}g_-)(x)g_+(x) &\le (\frac{d}{dx}g_-)(0)g_+(r_1) + \lambda g_-(r_1)\limsup_{x\uparrow r_1} m(x)g_+(x) \\
&= \lambda g_-(r_1)\limsup_{x\uparrow r_1} m(x)g_+(x) \\
&= 0
\end{aligned}$$

if r_1 is accessible. For, in this case $g_-(r_1) < \infty$, $m(r_1) < \infty$, $g_+(r_1) = 0$. Thus, if r_1 is accessible, then $(Gf)(r_1) = 0$ in case f vanishes in a neighborhood of r_1 or if $f \equiv 1$.

Any $f \in C(I)$ may be approximated uniformly by a linear combination of functions of these two types. Also (21.50) implies $(W(G1)(x) \le \lambda^{-1}W$ for all $x)$

$$||Gf||_\infty \le \lambda^{-1}||f||_\infty, \tag{21.52}$$

which holds whatever the type of boundaries. Therefore, the assertion concerning an *accessible boundary* is proved.

If r_1 is *natural*, and $f \equiv 1$, then $(\frac{d}{dx}g_+)(r_1) = 0$ and $g_+(r_1) = 0$ by (21.50),

$$(Gf)(x) = \lambda^{-1} - \frac{\lambda^{-1}}{W}g_+(x)(\frac{d}{dx}g_-)(r_0) \to \lambda^{-1} \text{ as } x \uparrow r_1, \tag{21.53}$$

while, as shown earlier, $Gf(x) \to 0$ as $x \uparrow r_1$ for all f vanishing in a neighborhood of r_1. Therefore, using (21.52) and approximation,

$$\lim_{x\uparrow r_1}(Gf)(x) = \lambda^{-1}f(r_1) \tag{21.54}$$

for every $f \in C(I)$, in case r_1 is natural.

It remains to consider the case: r_1 *is entrance*. In this case $(\frac{d}{dx}g_+)(r_1) = 0$, and (21.50) yields

$$\begin{aligned}
(G1)(x) &= \lambda^{-1} - \frac{\lambda^{-1}}{W}g_+(x)(\frac{d}{dx}g_-)(r_0) \\
&\to \lambda^{-1} - \lambda^{-1}g_+(r_1)(\frac{d}{dx}g_-)(r_0) = \lambda^{-1} \text{ as } x \to r_1.
\end{aligned} \tag{21.55}$$

■

We are now in a position to complete the proof of Theorem 21.1.

Proof of Theorem 21.1

(a) (*Sufficiency*) $\mathcal{D}_L$ is dense in $C(I)$ by Lemma 1. Fix $\lambda > 0$. By Lemmas 7 and 8, $u \equiv Gf \in \mathcal{D}_L$ and solves

$$(\lambda - L)u = f. \tag{21.56}$$

for every $f \in C(I)$. Fix $f \in C(I)$. Let us show that $u \equiv Gf$ is the unique element of $\mathcal{D}_L$ satisfying (21.56). If u_1 is another such element, then $(\lambda - L)(u - u_1)(x) = 0$ for $r_0 < x < r_1$. Therefore by Lemma 3, $u - u_1 = c_1 g_+ + c_2 g_-$ for some constants c_1, c_2. If r_1 is inaccessible, then $g_-(r_1) = \infty$ (Lemma 6 (III)) and $g_+(r_1) < \infty$. Therefore in order that $u - u_1 \in C(I)$, one must have $c_2 = 0$. Similarly, the inaccessibility of r_0 would imply $c_1 = 0$ in order that $u - u_1 \in C(I)$. Hence $u - u_1 \equiv 0$. Thus $u = Gf$ is the unique element of $\mathcal{D}_L$ satisfying (21.56), i.e., $\lambda - L$ is one-to-one on $\mathcal{D}_L$ and its range is $C(I)$. Further, (21.52) implies that

$$||(\lambda - L)^{-1} f||_\infty = ||Gf||_\infty \le \lambda^{-1} ||f||_\infty.$$

Thus, by the Hille–Yosida Theorem A.1, L on $\mathcal{D}_L$ is the infinitesimal generator of a strongly continuous contraction semigroup on $C(I)$.
(*Necessity*) If r_1 is accessible, then $0 < g_-(r_1) < \infty$. Hence for any given $f \in C(I)$, $Gf + c_1 g_- \in \mathcal{D}_L$ and is a solution of $(\lambda - L)u = f$ for every $c_1 \in \mathbb{R}$. Thus $\lambda - L$ is not one-to-one. Similarly, if r_0 is accessible then $Gf + c_2 g_+ \in \mathcal{D}_L$ and is a solution of $(\lambda - L)u = f$ for every $c_2 \in \mathbb{R}$. Thus if r_0, r_1 are not both inaccessible, L on $\mathcal{D}_L$ is not the infinitesimal generator of a strongly continuous semigroup on $C(I)$.

(b) Let T_t denote the semigroup generated by L (with domain $\mathcal{D}_L$) on $C(I)$. By the Riesz representation theorem, there exists a finite signed measure $p(t; x, dy)$ such that

$$(T_t f)(x) = \int_I f(y) p(t; x, dy) \quad \text{for all } f \in C(I). \tag{21.57}$$

The Laplace transform of $t \to (T_t f)(x)$ is the resolvent (see (2.47), Chapter 2)

$$(\lambda - L)^{-1} f(x) = \int_0^\infty e^{-\lambda t} (T_t f)(x) dt \quad (t \ge 0). \tag{21.58}$$

By uniqueness of the solution of $(L - \lambda)u = f$,

$$(\lambda - L)^{-1} f(x) = (Gf)(x). \tag{21.59}$$

But for $f \ge 0$, $(Gf)(x) \ge 0$. This implies that $p(t; x, dy)$ is a measure, and (21.52) implies that $0 \le p(t; x, B) \le 1$ for all Borel subsets B of I. Measurability of $x \to p(t; x, B)$ follows from that of $x \to T_t f(x)$ for all $f \in C(I)$ (Exercise 2(i)). The

Chapman–Kolmogorov equation is just a restatement of the semigroup property of T_t. Also, since r_0, r_1 are inaccessible $(\frac{dg_+}{dx})(r_1) = 0$, $(\frac{dg_-}{dx})(r_0) = 0$. Hence (21.50) yields

$$(G1)(x) = \frac{1}{\lambda} \quad \text{for all } x \in I. \tag{21.60}$$

Since the Laplace transform is one-to-one on the set of all bounded continuous functions on $[0, \infty)$ (Exercise 2(ii)), and

$$\frac{1}{\lambda} = \int_0^\infty e^{-\lambda t} dt, \tag{21.61}$$

it follows that

$$p(t; x, I) \equiv \int_I 1 p(t; x, dy) = 1, \text{ (for all } t > 0, x \in I). \tag{21.62}$$

Thus $p(t; x, dy)$ is a transition probability function on I. Feller continuity is simply the statement $x \to T_t f(x)$ is continuous if $f \in C(I)$, which is clearly true.

It remains to prove (21.14). Given $\varepsilon > 0$ and $x \in I$, there exists $f_x \in \mathcal{D}_L$ such that (Exercise 4), $f_x(y) \ge 0$ for all $y \in I$ and

$$f_x(y) = \begin{cases} 0 & \text{for } |y - x| \le \frac{\varepsilon}{3}, \\ 1 & \text{for } |y - x| \ge \frac{2\varepsilon}{3}. \end{cases} \tag{21.63}$$

Then if $|x' - x| \le \varepsilon/3$, one has $f_x(x') = 0$ and

$$\begin{aligned} &\sup_{\{x' \in I : |x'-x| < \frac{\varepsilon}{3}\}} \frac{T_t f_x(x') - f_x(x')}{t} \\ &= \sup_{\{x' \in I : |x'-x| < \frac{\varepsilon}{3}\}} \frac{T_t f_x(x')}{t} \\ &\ge \sup_{\{x' \in I : |x'-x| < \frac{\varepsilon}{3}\}} \frac{1}{t} \int_{\{y \in I : |y-x| \ge \frac{2\varepsilon}{3}\}} f(y) p(t; x', dy) \\ &= \sup_{\{x' \in I : |x'-x| < \frac{\varepsilon}{3}\}} \frac{1}{t} p(t; x', \{y : |y - x| \ge \frac{2\varepsilon}{3}\}) \\ &\ge \sup_{\{x' \in I : |x'-x| < \frac{\varepsilon}{3}\}} \frac{1}{t} p(t; x', \{y : |y - x'| \ge \varepsilon\}). \end{aligned} \tag{21.64}$$

But $Lf_x(x') = 0$ if $|x' - x| \le \frac{\varepsilon}{3}$, and

$$\sup_{x' \in I} \left| \frac{T_t f_x(x') - f_x(x')}{t} - L f_x(x') \right| \to 0 \text{ as } t \downarrow 0. \tag{21.65}$$

Hence

$$\sup_{\{x' \in I : |x' - x" \leq \frac{\varepsilon}{3}\}} \frac{1}{t} p(t; x', \{y - x'| \geq \varepsilon\}) \to 0 \text{ as } t \downarrow 0. \tag{21.66}$$

Assume without essential loss of generality that I is a finite interval. This is achieved, e.g., by taking as a diffeomorphic image of (r_0, r_1) a finite interval (a, b) and then treating $[a, b]$ as I. This may mean abandoning the natural scale; but that is of no consequence. (In any case, the two-point compactification reduces I to a compact set). Now find a finite set of points $r_0 = x_0 < x_1 < \cdots < x_n = r_1$ such that $|x_{i-1} - x_i| \leq \frac{\varepsilon}{3}$ $(i = 0, 1, \ldots, n-1)$. Obtain nonnegative functions f_{x_i} with the properties (21.63) (with $x_i = x$). Then (21.66) holds with $x = x_i$ $(i = 1, 2, \ldots, n)$. Hence,

$$\begin{aligned} \sup_{x \in I} \frac{1}{t} p(t; x, \{y : |y - x| \geq \varepsilon\}) &= \max_{0 \leq i \leq n} \sup_{x' \in I : |x' - x_i| \leq \frac{\varepsilon}{3}} \frac{1}{t} p(t; x', \{y \in I : |y - x'| \geq \varepsilon\}) \\ &\to 0 \text{ as } t \downarrow 0. \end{aligned}$$

The proof of Theorem 21.1 is now complete. ■

The "necessity" part of Theorem 21.1 (and its proof) shows that if at least one of the boundary points r_i $(i = 0, 1)$ is accessible, then there exist more than one solution $u \in \mathcal{D}_L$ to $(\lambda - L)u = f$ for any given $f \in C(I)$. Thus in order to construct a semigroup in this case, it is necessary to restrict the domain of L further (by means of what may be called "boundary conditions"). The next two theorems precisely tell us these restrictions. In preparation we will require the processes associated with the generator L in Theorem 21.1.

For the statement of the following theorem, consider the space $S^{[0,\infty)}$ of all functions on $[0, \infty)$ into a Polish space (S, ρ), with the Kolmogorov σ-field $\mathcal{G}$ generated by finite or countably many coordinates. Denote by $X(t)$ the projection on the t-th coordinate: $X(t)(\omega) = \omega(t), t \geq 0, \omega \in S^{[0,\infty)}$. Let P_x denote the distribution on $(S^{[0,\infty)}, \mathcal{G})$ of the Markov process with a transition probability $p(t; y, dz)$, starting at $X(0) = x$.

Theorem 21.2 (Dynkin-Snell Criterion for Continuity of Sample Paths of Markov Processes) *Assume that for every* $\varepsilon > 0$,

$$\lim_{t \downarrow 0} \frac{1}{t} \sup_{x \in S} p(t; x, B^c(x : \varepsilon)) = 0, \tag{21.67}$$

where

$$B(x : \varepsilon) = \{y \in S : \rho(y, x) < \varepsilon\}, \quad B^c(x : \varepsilon) = S \backslash B(x : \varepsilon).$$

Let $\Omega = C([0,\infty) : S)$ be the set of all continuous functions on $[0,\infty)$ into S, considered as a subset of the product space $S^{[0,\infty)}$. Then

$$P_x^*(\Omega) = 1, \tag{21.68}$$

where P_x^* denotes the outer measure on subsets of $S^{[0,\infty)}$ defined by the probability space $(S^{[0,\infty)}, \mathcal{G}, P_x)$.

For the proof of the theorem, we will need the following lemma.

Lemma 9 *For $\varepsilon > 0$, $\delta > 0$, define*

$$\Gamma(\varepsilon,\delta) = \sup_{t<\delta}\sup_{x\in S} p(t; x, B^c(x:\varepsilon)). \tag{21.69}$$

Let D be a countable subset of $[t_0, t_1]$, $0 \le t_0 < t_1 < \infty$. Then,

(a) $P_x(\cup_{s,t\in D}[\rho(X(s), X(t)) \ge 4\varepsilon]) \le 2\Gamma(\varepsilon, t_1 - t_0)$.
(b) $P_x(\cup_{s,t\in D,|t-s|<\delta}[\rho(X(s), X(t)) \ge 4\varepsilon]) \le 2((t_1 - t_0)/\delta)\Gamma(\varepsilon,\delta)$.

Proof First assume D is a finite set,

$$D = \{s_1, s_2, \ldots, s_m\} \subset [t_0, t_1], \quad s_1 < s_2 < \cdots < s_m.$$

Then

$$A \equiv \cup_{s,t\in D}[\rho(X(s), X(t)) \ge 4\varepsilon] \subset A_0 \cup \cup_{1\le i\le m} A_i,$$

where, for $i = 1,\ldots,m$, $A_0 = [\rho(X(t_1), X(t_0)) \ge \varepsilon]$,

$$A_i := [\rho(X(t_0), X(s_j)) < 2\varepsilon, 1 \le j < i, \rho(X(t_0), X(s_i)) \ge 2\varepsilon, \rho(X(t_1), X((s_i)) \ge \varepsilon]. \tag{21.70}$$

To verify (21.70) note that, if there exist $s, t \in D$ such that

$$\rho(\omega(s), \omega(t)) \ge 4\varepsilon,$$

then there exists a smallest i such that

$$\rho(\omega(t_0), \omega(s_i)) \ge 2\varepsilon.$$

If, in addition, $\rho(\omega(t_1), \omega(t_0)) < \varepsilon$, then one must have $\rho(\omega(t_1), \omega(s_i)) \ge \varepsilon$. By définition of Γ,

$$P_x(A_0) \le \Gamma(\varepsilon, t_1 - t_0), \tag{21.71}$$

and, by the Markov property, writing

$$C_i = \{(y_0, \ldots, y_i) \in S^{i+1} : \rho(y_0, y_j) < 2\varepsilon, 1 \le j < i, \rho(y_0, y_i) \ge 2\varepsilon\},$$

one has for $i = 1, \ldots, m$,

$$\begin{aligned} P_x(A_i) &= \mathbb{E}_x(\mathbf{1}[(X(t_0), X(s_1), \ldots, X(s_i)) \in C_i] P_y(\rho(X(t_1 - s_i), y) \ge \varepsilon)_{y=X(s_i)} \\ &\le \mathbb{E}_x(\mathbf{1}[(X(t_0), X(s_1), \ldots X(s_i)) \in C_i])\Gamma(\varepsilon, t_1 - t_0). \end{aligned} \tag{21.72}$$

Since the events F_i, say, within the brackets in the last expectation are disjoint, one has

$$P_x(\cup_{1 \le i \le m} A_i) \le \Gamma(\varepsilon, t_1 - t_0) P_x(\cup_{1 \le i \le m} F_i) \le \Gamma(\varepsilon, t_1 - t_0). \tag{21.73}$$

Combining (21.71) and (21.73), one gets (a) for a finite set D. Observe that the right side of the inequality in (a) does not depend on the location or number of points. Since the set of first n points in an enumeration of D increases with n, the assertion (a) is thus proved for a denumerable set D. Inequality (b) follows from (a) on dividing up $[t_0, t_1]$ into subintervals of length δ, and applying (a) to points in the union of each pair of adjacent intervals. ■

Proof of the Dynkin-Snell Theorem 21.2 To prove the theorem, let $A \in \mathcal{G}$ be such that $\Omega \subset A$. We need to show $P_x(A) = 1$. There exists a countable set $D \subset [0, \infty)$ such that A is (measurably) determined by coordinates in D. Without loss of generality (adding points to D if necessary), assume that D is dense in $[0, \infty)$. Let

$$D_m = D \cap [m - 1, m + 1], \quad (m = 1, 2, \ldots).$$

Fix $\eta > 0$. Find $0 < \delta(m, n) \downarrow 0$ as $(m, n) \uparrow$ to satisfy

$$(2/\delta(m, n))\Gamma(1/n, \delta(m, n)) \le 2^{-(m+n)}\eta/2m \quad (m, n \ge 1). \tag{21.74}$$

This is possible by hypothesis (21.67); for instance, take $\varepsilon = 1/n, \delta(m, n)$ sufficiently small with m sufficiently large, depending on n, and $t < \delta(m, n)/2$. By part (b) of the lemma,

$$P_x(\cup_{s,t \in D_m, |t-s| < \delta(m,n)/2}[\rho(X(s), X(t)) \ge 4/n)]) \le 2^{-(m+n)}\eta.$$

Taking the complement of the event within brackets above, one obtains

$$P_x(\cap_{m \ge 1, n \ge 1, s,t \in D_m, |t-s| < \delta(m,n)/2}[\rho(X(s), X(t)) < 4/n]) \ge 1 - \eta. \tag{21.75}$$

Now, since each element of the event

$$F := \cap_{m \ge 1, n \ge 1, s,t \in D_m, |t-s| < \delta(m,n)/2}[\rho(X(s), X(t)) < 4/n]$$

is uniformly continuous on D_m for every $m \geq 1$, one has

$$F \subset \tilde{A} := \{\omega \in S^{[0,\infty)} : \omega \text{ is uniformly continuous on } D_m \ \forall\, m \geq 1\} \in \mathcal{G}.$$

It follows that

$$P_x(\tilde{A}) \geq 1 - \eta, \quad \forall\, \eta > 0,$$

and hence,

$$P_x(\tilde{A}) = 1.$$

On the other hand, $\tilde{A} \subset A$. To see this, let $\tilde{\omega} \in \tilde{A}$. There is a unique extension of $\tilde{\omega}$ to an element $\omega_0 \in \Omega = C([0,\infty) : S)$ such that $\omega_0(t) = \tilde{\omega}(t)\ \forall\, t \in D$. Since $\omega_0 \in A$, $\tilde{\omega} \in A$. For the coordinates of elements of A outside of D are free and, in particular, include $\tilde{\omega}$. Therefore, $P_x(A) = 1$. ■

We have the following immediate corollary.

Corollary 21.3 *Under the hypothesis of Theorem 21.1, with probability one, the Markov process has continuous sample paths.*

Remark 21.4 Under relaxed conditions along the same lines, the methods in proving Theorem 21.2. may be modified to achieve right-continuous paths with left-hand limits, i.e., so-called cadlag paths. A precise statement is as follows. The proof is left as an exercise.

Theorem 21.4 (Dynkin Criterion[6] for Right Continuity of Sample Paths of Markov Processes) *In place of (21.67), assume that for any $\varepsilon > 0$,*

$$\lim_{t\downarrow 0} \sup_{x \in S} p(t; x, B^c(x : \varepsilon)) = 0. \tag{21.76}$$

Then, for a probability measure μ on $(S, \mathcal{B})$, there is a Markov process $\{X_t : t \geq 0\}$ with transition probabilities $p(t; x, dy)$ and initial distribution μ, defined on a probability space $(\Omega, \mathcal{F}, P_\mu)$ having right-continuous sample paths with left limits.

Theorem 21.5 *Let $p(t; x, dy)$ be a Feller continuous (sub)-probability on $I = [r_0, r_1]$ such that the infinitesimal generator A of the corresponding semigroup on $C(I)$ has domain $\mathcal{D}_A \subset \mathcal{D}_L$, and $(Af)(x) = (Lf)(x)$ for all $x \in (r_0, r_1)$ and every $f \in \mathcal{D}_A$. If r_0, r_1 are regular boundary points, then every $f \in \mathcal{D}_A$ satisfies "boundary conditions" of the form*

[6] Dynkin (1965), Vol 1, Theorem 3.6, p.92.

$$\Phi_i(f) = \alpha_i f(r_i) + \int \frac{f(r_i) - f(x)}{|r_i - x|} dQ_i(x) + \beta_i (Lf)(r_i) = 0 \quad (i = 0, 1) \tag{21.77}$$

where α_i, β_i are nonnegative real numbers, r_i is an extended real number, Q_i is a finite measure on I and, either

$$\beta_i > 0 \text{ or } \int_I \frac{dQ_i(x)}{|r_i - x|} = \infty \tag{21.78}$$

Remark 21.5 The integrand in (21.77) is taken to be $(-1)^{i+1} f'(r_i^{\pm})$ for $x = r_i$, respectively, for $i = 0, 1$.

Proof Write

$$\begin{aligned} r(t) &= 1 - p(t; r_0, I), \quad q(t) = \int_I (y - r_0) p(t; r_0, dy), \\ \pi(t; dy) &= \frac{(y - r_0) p(t; r_0, dy)}{q(t)}. \end{aligned} \tag{21.79}$$

Note that $0 \leq r(t) \leq 1$, $0 \leq q(t) \leq r_1 - r_0 < \infty$.

For $f \in \mathscr{D}_A$ one has, noting $p(t; r_0, I) = 1 - r(t)$, and that $I = [r_0, r_1]$ is compact,

$$\begin{aligned} (Af)(r_0) &= \lim_{t \downarrow 0} \frac{1}{t} \{T_t f(r_0) - f(r_0)\} \\ &= \lim_{t \downarrow 0} \frac{1}{t} [\int_I (f(y) - f(r_0)) p(t; r_0, dy) - r(t) f(r_0)] \\ &= \lim_{t \downarrow 0} \frac{1}{t} [q(t) \int_I \frac{f(y) - f(r_0)}{y - r_0} \pi(t; dy) - r(t) f(r_0)]. \end{aligned} \tag{21.80}$$

Choose a sequence $t_k \downarrow 0$ such that the following limits exist as extended real numbers:

$$\begin{gathered} \tfrac{q(t_k)}{t_k} \to Q, \quad \tfrac{r(t_k)}{t_k} \to R, \quad \tfrac{q(t_k)}{r(t_k)} \to S, \\ \pi(t_k; dy) \Rightarrow \pi_0(dy) \text{ (weakly)}, \end{gathered} \tag{21.81}$$

where Q, R, S may be infinite and π_0 is a probability measure on I. First assume $Q < \infty$, $R < \infty$. Then

$$(Af)(r_0) = Q \int_I \frac{f(y) - f(r_0)}{y - r_0} \pi_0(dy) - Rf(r_0). \tag{21.82}$$

Since $(Af)(y) = (Lf)(y)$ for $y > r_0$ and $Af \in C(I)$

$$(Af)(r_0) = \lim_{y \downarrow r_0} (Af)(y) = (Lf)(r_0), \tag{21.83}$$

which is of the form (21.77) for $i = 0$, with

$$\alpha_0 = R, \;\; Q_0(dy) = AQ\pi_0(dy), \;\; \beta_0 = 1. \tag{21.84}$$

In case $Q + R = \infty$, one has

$$\begin{aligned}
0 = \frac{(Af)(r_0)}{Q+R} &= \lim_{t_k \downarrow 0} \frac{\frac{1}{t_k}[q(t_k)\int_I \frac{(r(y)-f(r_0))}{y-r_0}\pi(t_k; dy) - r(t_k)f(r_0)]}{\frac{1}{t_k}(q(t_k)+r(t_k))} \\
&= \frac{\int_I \frac{f(y)-f(r_0)}{y-r_0}\pi_0(dy)}{1+\frac{1}{S}} - \frac{f(r_0)}{1+S} \\
&= \frac{S}{1+S}\int_I \frac{f(y)-f(r_0)}{y-r_0}\pi_0(dy) - \frac{1}{1+S}f(r_0),
\end{aligned} \tag{21.85}$$

which is of the form (21.77) for $i = 0$. If $S = \infty$ then one must take $\frac{S}{S+1} = 1$ in the above computation and (21.85) becomes

$$\int_I \frac{f(y)-f(r_0)}{y-r_0}\pi_0(dy) = 0, \tag{21.86}$$

and if $S = 0$, (21.85) becomes $f(r_0) = 0$. In any case one has a relation of the form $\Phi_0(f) = 0$, with $\beta_0 = 0$ and at least one of α_0, Q_0 not zero.

Finally if $\beta_i = 0$ and $\int \frac{dQ_i(x)}{|r_i - x|} < \infty$, then $f \to \Phi_i(f)$ is a nonzero bounded linear functional on $C(I)$ (indeed $|\Phi_i(f)| \leq d_i||f|| + 2||f||\left(\int \frac{dQ_i(x)}{|r_i-x|}\right)$). Hence $\{f \in C(I) : \Phi_i(f) = 0\}$ is a *closed* linear subspace of $C(I)$ which is *not* equal to $C(I)$ and is, therefore, not dense in $C(I)$. As a consequence $\mathscr{D}_A = \mathcal{D}_L \cap \{f : \Phi_i(f) = 0\}$ is not dense in $C(I)$ contradicting part (b) of the theorem on p. 6. ■

We now show that each pair of boundary conditions

$$\Phi_i(f) = 0, \quad i = 0, 1,$$

gives rise to a unique strongly continuous *Markov semigroup* $\{T_t\}$ on $C(I)$: i.e., the measure $p(t; x, dy)$ in the Riesz representation of the bounded linear functional $f \to T_t f(x)$ is a transition (sub)-probability on I for each fixed $t \geq 0$, $x \in I$.

Theorem 21.6 *Let $\Phi_i(f) = 0$ ($i = 0, 1$) be nontrivial boundary conditions of the form (21.77) which satisfy (21.78). Let A be the operator defined on*

$$\mathscr{D}_A = \mathcal{D}_L \cap \{f : \Phi_i(f) = 0 \, for \, i = 0, 1\}$$

by

$$(Af)(x) = (Lf)(x), r_0 < x < r_1.$$

Then A is the infinitesimal generator of a strongly continuous contraction semigroup on $C(I)$ or a proper closed subspace of $C(I)$.

Proof The proof will be carried out in several steps. First let us normalize the functions g_+, g_- (see (21.13), (21.20)). Write

$$v_+ = \frac{g_+}{g_+(r_0)}, \quad v_- = \frac{g_-}{g_-(r_1)}. \tag{21.87}$$

Then

$$v_+(r_0) = 1,\ v_+(r_1) = 0;\ v_-(r_0) = 0,\ v_-(r_1) = 1. \tag{21.88}$$

Choose and fix $\lambda > 0$ arbitrarily. Given $f \in C(I)$ the general solution of

$$(\lambda - L)u(x) = f(x), \quad r_0 < x < r_1, \tag{21.89}$$

is (see Lemma 3)

$$u = Gf + c_1 v_+ + c_2 v_- \tag{21.90}$$

where Gf is defined by (21.45), for suitable constants $c_i \equiv c_i(f)$. The boundary conditions for u may be expressed as

$$\begin{bmatrix} \Phi_0(v_+) & \Phi_0(v_-) \\ \Phi_1(v_+) & \Phi_1(v_-) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = - \begin{bmatrix} \Phi_0(Gf) \\ \Phi_1(Gf) \end{bmatrix}. \tag{21.91}$$

We need to show that the 2×2 matrix on the left is nonsingular.

Consider the function

$$v(x) = v_+(x) + v_-(x). \tag{21.92}$$

It satisfies: (i) $D_m D_x^+ v = \lambda v > 0$, (ii) $v(r_0) = v(r_1) = 1$. It follows from the strictly increasing property of $D_x^+ v$ (implied by (i)) that v is strictly convex and, hence, (a) $v(x) < 1$ for $r_0 < x < r_1$, and (b) $D_x^+ v(r_0) < 0$, $D_x^+ v(r_1) > 0$, i.e., $D_x^+ v_+(r_0) < -D_x^+ v_-(r_0)$, $D_x^+ v_+(r_1) > -D_x^+ v_-(r_1)$.

Now

$$\Phi_0(v_+) = (\alpha_0 + \beta_0) + \int_I \frac{1 - v_+(x)}{x - r_0} dQ_0(x),$$
$$\Phi_1(v_+) = \int_I \frac{-v_+(x)}{r_1 - x} dQ_1(x),$$
$$\Phi_0(v_-) = \int_I \frac{-v_-(x)}{x - r_0} dQ_0(x),$$
$$\Phi_1(v_-) = (\alpha_1 + \beta_1) + \int_I \frac{1 - v_-(x)}{r_1 - x} dQ_1(x), \tag{21.93}$$

By property (a) above, $\frac{1-v_+(x)}{x-r_0} > |\frac{-v_-(x)}{x-r_0}| = \frac{v_-(x)}{x-r_0}$ for all $x \neq r_0$; since $\alpha_0+\beta_0 \geq 0$, it then follows that

$$\Phi_0(v_+) > |\Phi_0(v_-)| \tag{21.94}$$

unless

$$\alpha_0 = \beta_0 = 0 \text{ and } Q_0(I - \{r_0\}) = 0. \tag{21.95}$$

Similarly,

$$\Phi_1(v_-) > |\Phi_1(v_+)| \tag{21.96}$$

unless

$$\alpha_1 = \beta_1 \text{ and } Q_1(I - \{r_1\}) = 0. \tag{21.97}$$

If $\alpha_0 = \beta_0 = 0$ and $Q_0(I - \{r_0\}) = 0$, then $Q_0(\{r_0\}) \equiv q_0 > 0$, and

$$\Phi_0(v_+) = -q_0(D_x^+ v_+)(r_0),\ \Phi_0(v_-) = -q_0(D_x^+ v_-)(r_0),$$

and, by property (b) above,

$$\Phi_0(v_+) > |\Phi_0(v_-)|. \tag{21.98}$$

Similarly, if $\alpha_1 = \beta_1 = 0$ and $Q_1(I - \{r_1\}) = 0$, then

$$\Phi_1(v_-) > |\Phi_1(v_1)|. \tag{21.99}$$

Thus in *all* cases, the determinant of the 2×2 matrix in (21.91) is positive, so that c_1, c_2 are *uniquely* determined by the boundary conditions (21.91). Hence there exists a unique element $u \in \mathscr{D}_A$ such that

$$(\lambda - A)u = f. \tag{21.100}$$

Next let us show that

$$f \geq 0 \text{ implies } u \geq 0,$$

where u is the solution in $\mathscr{D}_A$ of (21.100). For this note that (see Lemma 8)

$$(Gf)(r_i) = 0, \quad i = 0, 1,$$

and, (see the proof of Lemma 8 and Table 21.1),

$$(D_x^+ Gf)(r_0) \geq 0, \text{ and } (D_x^+ Gf)(r_1) \leq 0.$$

Hence,

$$\begin{aligned} \Phi_0(Gf) &= \textstyle\int_I \frac{-(Gf)(x)}{x - r_0} dQ_0(x) \leq 0, \\ \Phi_1(Gf) &= \textstyle\int_I \frac{-(Gf)(x)}{r_1 - x} dQ_1(x) \leq 0. \end{aligned} \tag{21.101}$$

Now the diagonal elements of the matrix in (21.91) are positive, off-diagonal elements nonpositive, and the determinant of the matrix is positive (by calculations (21.93–(21.99)). Therefore, the elements of the inverse matrix are all nonnegative. It then follows from (21.91) and (21.101) that $c_0 \geq 0$, $c_1 \geq 0$. Therefore, $u(x) = (Gf)(x) + c_1 v_+(x) + c_2 v_-(x) \geq 0$ for all x.

Finally, one needs to show that the solution u in $\mathscr{D}_A$ of

$$(\lambda - A)u = f$$

satisfies, for each $f \in C(I)$, the inequality

$$||u|| \leq \frac{1}{\lambda} ||f||. \tag{21.102}$$

Because of the fact that $f_1 \geq f_2$ implies $u_1 \geq u_2$, where u_i is the solution in $\mathscr{D}_A$ of

$$(\lambda - A)u_i = f_i \quad i = 1, 2,$$

in order to prove (21.102), it is enough to prove (21.102) for $f \equiv 1$. Since the function

$$u(x) = \lambda^{-1} + k_1 v_+(x) + k_2 v_-(x) \tag{21.103}$$

is in $\mathcal{D}_L$ and satisfies $(\lambda - L)u(x) = 0$ $(r_0 < x < r_1)$, whatever be the constants k_1, k_2, the unique solution to

$$(\lambda - A)u = \lambda^{-1}$$

in $\mathscr{D}_A$ is given by (21.103) with k_1, k_2 determined by the equations

$$0 = \Phi_0(\lambda^{-1}) + k_1\Phi_0(v_+) + k_2\Phi_0(v_-)$$
$$0 = \Phi_1(\lambda^{-1}) + k_1\Phi_1(v_+) + k_2\Phi_1(v_-),$$

i.e., (writing Δ for the determinant of the 2×2 matrix discussed earlier),

$$\begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \frac{1}{\Delta}\begin{bmatrix} \Phi_1(v_-) & -\Phi_1(v_+) \\ -\Phi_0(v_-) & \Phi_0(v_+) \end{bmatrix}\begin{bmatrix} -\Phi_0(\lambda^{-1}) \\ -\Phi_1(\lambda^{-1}) \end{bmatrix}. \tag{21.104}$$

But $\Phi_0(\lambda^{-1}) = (\alpha_0 + \beta_0)\lambda^{-1} \geq 0$, $\Phi_1(\lambda^{-1}) = (\alpha_1 + \beta_1)\lambda^{-1} \geq 0$. Hence $k_1 \leq 0$, $k_2 \leq 0$ (note that the elements of the 2×2 matrix in (21.104) are all nonnegative). Thus $u(x)$ appearing in (21.103), with k_1, k_2 given by (21.104), satisfies

$$u(x) \leq \lambda^{-1} \text{ for } r_0 \leq x \leq r_1. \tag{21.105}$$

To complete the proof of the theorem, one still needs to prove that $\mathscr{D}_A$ is dense in $C(I)$. However, this follows from the following general fact from functional analysis (Exercise 18).

Lemma 10 *If Φ is an unbounded linear functional on a Banach space $\mathscr{X}$, then its kernel*

$$\{f \in \mathscr{X} : \Phi(f) = 0\}$$

is dense in $\mathscr{X}$. In fact, if Φ_i $(1 \leq i \leq k)$ are a finite number of unbounded linear functionals, then the set $\{f \in \mathscr{X} : \Phi_i(f) = 0\ \forall\, i\}$ is dense in $\mathscr{X}$

The proof of this lemma is outlined in Exercise 18. ■

Remark 21.6 In the case that one boundary, say r_0, is accessible, and the other r_1 is inaccessible, then as noted in the proof of Theorem 21.1, for $f \in C(I)$, $(\lambda - D_m D_x^+)^{-1} f(x) = Gf(x) + cv_+(x)$ for some constant c. If both r_0 and r_1 are inaccessible then, again as shown in the proof of Theorem 21.1, the resolvent is Gf.

One may rewrite the boundary condition (21.77) as (see the note following the statement of Theorem 21.5):

$$\alpha_i f(r_i) + (-1)^{i+1}\gamma_i(D_x^+ f)(r_i) + \int_{I\setminus\{r_i\}} \frac{f(r_i) - f(x)}{|r_i - x|} dQ_i(x) + \beta_i(Lf)(r_i) = 0, \tag{21.106}$$

where $\gamma_i = Q_i(\{r_i\})$.

Definition 21.7 The boundary condition (21.106) is said to be *local* if the integral is zero, i.e., if $Q(I - \{r_i\}) = 0$.

The reason for this nomenclature is clear: a local boundary condition does not restrict the functions away from an arbitrarily small neighborhood of the boundary point.

Definition 21.8 The diffusion is *conservative* if $p(t; x, I) = 1$ for all $t > 0$ and all $x \in I$.

If $p(t; x, I) < 1$, then one may add a fictitious state r_∞, say, to the state space and obtain a transition probability (Exercise) on $\overline{I} = I \cup \{r_\infty\}$:

$$\begin{aligned}
\overline{p}(t; x, B) &= p(t; x, B) \;\text{ if } x \in I,\, B \in \mathcal{B}(I),\\
\overline{p}(t; x, \{r_\infty\}) &= 1 - p(t; x, I) \;\text{ if } x \in I,\\
\overline{p}(t; r_\infty, \{r_\infty\}) &= 1.
\end{aligned} \tag{21.107}$$

There are many other ways of augmenting p so as to obtain a transition probability. No matter how this is done up to the time of leaving I, the process is determined by p. More precisely, if $\{X(t) : t \geq 0\}$ is the Markov process with such a Feller continuous transition probability $\overline{p}$ having right continuous sample paths, letting r_∞ be an added isolated point in $\overline{I}$, while giving I the usual topology, then p is the transition sub-probability corresponding to the semigroup:

$$T_t f(x) = \mathbb{E}_x\{f(X(t)) \cdot \mathbf{1}_{[\tau > t]}\}, \quad x \in I,\; f \text{ bounded and measurable.} \tag{21.108}$$

To conclude this chapter, we shall consider a number of illustrative and important examples of Feller's boundary conditions and their probabilisitic meanings. To this end, let r_0 be a regular boundary in Feller's classification (Theorem 21.5), with natural scale. We first consider the local boundary conditions where the behavior at r_0 depends entirely only on the behavior of the diffusion in its immediate vicinity. We assume that r_1 is either inaccessible or has a local boundary condition as well.

Example 2 (Dirichlet Boundary Condition) We begin with the so-called Dirichlet boundary condition at r_0, namely,

$$f(r_0) = 0, \quad \Phi_1(f) = 0, \tag{21.109}$$

where $\Phi_1(f) = 0$ is a local boundary condition at r_1. This is the case in the general condition (21.106) with $\alpha_0 > 0$, and all other terms equal to 0 for $i = 0$. Then the domain $\mathcal{D}_L$ of the generator $L = \frac{d}{dm}\frac{d}{dx}$ is contained in $C(I) \cap \{f \in C(I) : f(r_0) = 0\}$. For $f \in \mathcal{D}_L$ this implies

$$\frac{d}{dt} T_t f = L T_t f \in \mathcal{D}_L,$$

so that

$$T_t f(r_0) = 0 \text{ for all } t.$$

Since $\mathcal{D}_L$ contains functions which are positive except at the boundaries, the last condition, namely,

$$\int_I f(y)p(t; r_0, dy) = 0 = \lim_{x \to r_0} \int_I f(y)p(t; x, dy) = 0 \,\forall\, t,$$

implies that on reaching r_0 the diffusion $X = \{X_t : t \geq 0\}$ vanishes. In particular, $p(t; r_0, I) = 0$. The same argument applies to the condition $f(r_1) = 0$. In particular, if (21.109) holds, then the diffusion is not conservative: $p(t; x, I) \leq P_x(\tau > t)$, where τ is the first time the diffusion reaches r_0, with equality under certain boundary conditions at r_1.

Example 3 (Absorbing Boundary Condition) Next consider the boundary conditions

$$Lf(r_0) = 0, \tag{21.110}$$

with r_1 inaccessible or local, i.e., (21.106) with $\beta_0 > 0$ and all other terms being zero for $i = 0$. The boundary condition at r_0 implies that for $f \in \mathcal{D}_L$,

$$\frac{d}{dt} T_t f(r_0) = LT_t f(r_0) = 0.$$

Hence $t \to T_t f(r_0)$ is a constant, namely, $f(r_0)$. Therefore,

$$p(t; r_0, \{r_0\}) = 1 \quad \forall\, t \geq 0.$$

We will call this an *absorbing boundary condition*. Here the diffusion stays at r_0 forever after reaching it. Once again, the same applies to r_1 with boundary condition $Lf(r_1) = 0$. This terminology is, unfortunately, at variance with that used by Feller (1954), where the Dirichlet boundary condition in which the process vanishes, $f(r_0) = 0$, is called "absorbing", and the boundary condition in which the process remains at the boundary, $Lf(r_0) = 0$, is called "adhesive". Our only justification is that in all our previous work, we have used the present terminology.

Example 4 (Reflecting Boundary Condition) Another local boundary condition at r_0 is given by

$$f'(r_0^+) \equiv \frac{d}{dx} f(r_0^+) = 0, \quad \Phi_1(f) = 0, \tag{21.111}$$

i.e., in (21.106) take $\gamma_0 > 0$ and all other terms zero for $i = 0$. This is referred to as r_0 being a *reflecting boundary*. It arises if, in (21.77), Q_0 is the point mass $\delta_{r_0}(dx)$ at r_0, $\alpha_0 = 0$, $\beta_0 = 0$. The integrand in (21.77) in this case is defined to be $f'(r_0^+)$. Here the diffusion returns to $(r_0, r_1]$ immediately, and continuously, after reaching r_0. In Chapter 13 it is shown that such a diffusion may be achieved by

folding a symmetrically extended process around r_0. In particular, if r_0 is taken to be zero, without loss of generality, then a diffusion $\{X_t : t \geq 0\}$ on $[0, r_1]$ with the given scale and speed functions can be extended to $[-r_1, r_1]$ by adjusting the scale function so that $s(0) = 0$, and letting $s(-x) = -s(x)$, and $m(-x) = -m(x)$. By replacing $m(x)$ by $m(x^+)$, one can make the left-continuous m on $[-r_1, 0)$ right-continuous. In any case for continuous m no such adjustment is necessary. Denoting the diffusion with this extended coefficient as $\{Y_t : t \geq 0\}$, the stochastic process $\{|Y_t| : t \geq 0\}$ on $[0, r_1]$ is shown to be a diffusion with the reflecting boundary condition (21.111) at 0 (Exercise 7). The diffusion is conservative if r_1 is inaccessible. If both boundaries are reflecting, i.e.,

$$f'(r_0^+) = 0, \quad f'(r_1^-) = 0, \tag{21.112}$$

then by the method of Chapter 13, a pathwise construction may be provided, and the diffusion is conservative. The method of Chapter 13 allows one to provide useful representations of transition probabilities of the transformed diffusion in terms of the original one. In (21.106) this is the case if $\gamma_0 > 0, \gamma_1 > 0$ and all other terms are 0.

Example 5 (Mixed (or Elastic) Boundary Condition) Our last example is a a *mixed local boundary* or *elastic local boundary* condition at r_0 (corresponding to $\alpha_0 > 0, \gamma_0 > 0$ and all other terms zero for $i = 0$ in (21.106)), namely, with $\gamma = \frac{\alpha_0}{\gamma_0}$,

$$f'(r_0^+) - \gamma f(r_0) = 0, \quad \Phi_1(f) = 0. \tag{21.113}$$

To describe the probabilistic behavior, without loss of generality take $r_0 = 0$, and recall the local time $\ell(t, \omega)$ of a standard Brownian motion $\{B_t : t \geq 0\}$ at 0 is the same as the local time for the reflected Brownian motion at 0. Here ω is a sample path of the reflected Brownian motion.

Proposition 21.7 *Let $X = \{X_t = |B_t| : t \geq 0\}$ be the reflected Brownian motion. Consider the stochastic process:*

$$\zeta(t, \cdot) = \exp\{-2\gamma \ell(t, \cdot)\}, \quad t \geq 0.$$

Let r_∞ be an arbitrary symbol, referred to as a "cemetary" state, adjoined to the state space. Define the process Y on $[0, \infty) \cup \{r_\infty\}$ by

$$Y_t = \begin{cases} X_t & \text{for} \quad t < \zeta, \\ r_\infty & \text{if} \quad t \geq \zeta, \end{cases} \tag{21.114}$$

where, conditionally given the σ-field $\mathcal{F}_t$ generated by $\{X_s : 0 \leq s \leq t\}$,

$$P_x(\zeta > t|\mathcal{F}_t) = \begin{cases} 1 & \text{if} \quad t < \tau_0, \\ \exp\{-2\gamma \ell(t - \tau_0)\} & \text{if} \quad t \geq \tau_0. \end{cases} \tag{21.115}$$

Here

$$\tau_0 = \inf\{s \geq 0 : X_s = 0\}.$$

Then the process $Z_t := \{Y_t : t < \zeta\}$ *is a diffusion on* $[0, \infty)$ *satisfying the boundary condition (21.113) at* r_0.

We will postpone the rather long proof of this proposition until after we conclude our next example, namely, that of a nonlocal boundary condition.

Example 6 (A Nonlocal Boundary Condition) One may represent the boundary condition (21.106), at r_i as

$$\Phi_i(f) \equiv \alpha_i f(r_i) - \Theta_i \int_I f(x)\mu_i(dx) - (-1)^i \gamma_i f'(r_i) + \beta_i Lf(r_i) = 0, \ (i = 0, 1), \tag{21.116}$$

where we assume

$$\Theta_i = \int_{I\backslash\{r_i\}} |r_i - x|^{-1} Q_i(dx) \equiv \Theta_i \int_I \mu_i(dx) < \infty, \tag{21.117}$$

μ_i being a probability measure on $I\backslash\{r_i\}$. The term $f(r_i)\int_{I\backslash\{r_i\}} |r_i - x|^{-1} Q_i(dx)$ is absorbed in $\alpha_i f(r_i)$. Consider the case $i = 0$, with $\Phi_1(f) = 0$, a local boundary condition at r_i, or r_1 is inaccessible. The diffusion on reaching r_0 stays there for a time η, which is independent of the path up to the time of arrival at r_0 and is exponentially distributed with parameter α_0. After time η, it jumps to a point, say Z_1, in $(r_0, r_1]$ randomly, and independently of the past, according to the distribution μ_0. Then the process starts afresh from the point Z_1 in $(r_0, r_1]$, in case of non-absorption if $Z_1 = r_1$, until it reaches r_0 again and repeats this process. If r_1 is an absorbing boundary and $Z_1 = r_1$, then the process stays at r_1 thereafter. The Markovian nature of the process is established by the same argument as used for jump Markov processes in Chapter 4, or by the fact that the process is generated by a contraction semigroup. To check the boundary condition, note that for $f \in C(I)$,

$$T_t f(r_0) = f(r_0)P(\eta > t) + \alpha_0 \int_{(0,t]} e^{-\alpha_0 s} ds \int_{(r_0,r_1]} f(x)\mu_0(dx), \tag{21.118}$$

so that

$$\frac{d}{dt} T_t f(r_0) = -\alpha_0 f(r_0)e^{-\alpha_0 t} + \alpha_0 e^{-\alpha_0 t} \int_{(r_0,r_1]} f(x)\mu_0(dx), \tag{21.119}$$

and, as a consequence,

$$Lf(r_0) = -\alpha_0 f(r_0) + \alpha_0 \int_{(r_0,r_1]} f(x)\mu_0(dx), \tag{21.120}$$

which is the same as (21.116) for $i = 0$ with $\Theta_0 = \alpha_0, \beta_0 = 1$. If the boundary condition at r_1 is local or r_1 is inaccessible, then this Markov process has continuous sample paths in between jumps from r_0.

We now provide a proof of Proposition 21.7.

Proof of Proposition 21.7 Let B be a Borel subset of $[0, \infty)$. Then, writing $\mathcal{F}_s = \sigma\{X_u : 0 \le u \le s\}$ and X_s^+ as the after-s X process: $(X_s^+)_t = X_{s+t}$, and noting that

$$[\zeta > s + t] = [\zeta > s, \zeta(X_s^+) > t],$$

one has

$$P_y((X_s^+)_t \in B, \zeta > s, \zeta(X_s^+) > t|\mathcal{F}_s) = \mathbf{1}_{[\zeta > s]} P_{X_s}(X_t \in B, \zeta > t) = P_y(Z_t \in B)|_{y=Z_s}, \tag{21.121}$$

proving the Markov property of the diffusion. To check that it satisfies the boundary condition (21.113), we consider the resolvent (Laplace transform) of $t \to f(t, \cdot) = \mathbb{E}_\cdot f(Z_t) \in \mathcal{D}_L, t \ge 0$, satisfying the boundary conditions (21.113). That is,

$$\begin{aligned}
&\hat{f}(\lambda, x) \\
&= \int_{[0,\infty)} e^{-\lambda t} \mathbb{E}_x f(Z_t) dt \\
&= \mathbb{E}_x \{\int_{[0,\infty)} e^{-\lambda t} e^{-2\gamma \ell(t)} f(X_t) dt\} \\
&= \mathbb{E}_x \{\int_{[0,\tau_0]} e^{-\lambda t} f(X_t) dt\} + \mathbb{E}_x \int_{[\tau_0,\infty)} (e^{-\lambda t} e^{-2\gamma \ell(X_{\tau_0}^+)} f((X_{\tau_0}^+)_t) dt \\
&= \mathbb{E}_x \int_{[0,\infty)} e^{-\lambda t} f(X_t) dt - \mathbb{E}_x \int_{[\tau_0,\infty)} e^{-\lambda t} f(X_t) dt + \mathbb{E}_x \{e^{-\lambda \tau_0} \hat{f}(\lambda, 0)\} \\
&= \mathbb{E}_x \int_{[0,\infty)} e^{-\lambda t} f(X_t) dt - \mathbb{E}_x (e^{-\lambda \tau_0} \mathbb{E}_0 \int_{[0,\infty)} e^{-\lambda t} f(X_t) dt) + \mathbb{E}_x [e^{-\lambda \tau_0} \hat{f}(\lambda, 0) \\
&= \mathbb{E}_x (e^{-\lambda \tau_0} (\hat{f}(\lambda, 0) - \mathbb{E}_0 \int_{[0,\infty)} e^{-\lambda t} f(X_t) dt)) + \mathbb{E}_x \int_{[0,\infty)} e^{-\lambda t} f(X_t) dt.
\end{aligned} \tag{21.122}$$

Recall[7] that

$$\mathbb{E}_x e^{-\lambda \tau_0} = e^{-x\sqrt{2\lambda}},$$

so that (21.122) becomes

$$\hat{f}(\lambda, x) = e^{-x\sqrt{2\lambda}} (\hat{f}(\lambda, 0) - \mathbb{E}_0 \int_{[0,\infty)} e^{-\lambda t} f(X_t) dt) + \mathbb{E}_x \int_{[0,\infty)} e^{-\lambda t} f(X_t) dt \tag{21.123}$$

[7] Bhattacharya and Waymire (2021), p. 85–86, Proposition 7.15 and Remark 7.8.

Writing

$$g(x) = \hat{f}(\lambda, x),$$

one has

$$g(x) = (\lambda - L)f(x),$$

or,

$$(\lambda - L)g(x) = f(x),$$

where L is the generator of the diffusion $\{Z_t : t \geq 0\}$. The derivative of g at $x = 0$ yields, using the fact that the derivative of the last summand in (21.123) vanishes (which is the boundary condition for the reflecting Brownian motion $\{X_t\}$),

$$\frac{d}{dx}g(x)|_{x=0} = -\sqrt{2\lambda}g(0) + \sqrt{2\lambda}\mathbb{E}_0 \int_{[0,\infty)} e^{-\lambda t} f(X_t)dt. \tag{21.124}$$

Now the last summand on the right in (21.124) can be expressed as (Exercise 9)

$$\sqrt{2\lambda}\mathbb{E}_0 \int_{[0,\infty)} e^{-\lambda t} f(X_t)dt = \int_{[0,\infty)} e^{-\sqrt{2\lambda}y} f(y)dy. \tag{21.125}$$

We will now evaluate $g(0) = \hat{f}(\lambda, 0)$ and show that

$$\sqrt{2\lambda}g(0) = \{\sqrt{2\lambda}/(\gamma + \sqrt{2\lambda})\} \int_{[0,\infty)} e^{-\sqrt{2\lambda}y} f(y)dy. \tag{21.126}$$

Plugging this and the quantity in (21.125) in (21.124) yields the desired relation (21.113), namely,

$$g'(0) = \gamma g(0). \tag{21.127}$$

Since every function $g \in \mathcal{D}_L$ is of the form $(\lambda - L)^{-1} f$ for some f in the center of the semigroup, the proof of the proposition would be complete, once (21.126) is proved. To derive (21.126) we follow the method of Itô-McKean (1965). Consider the alternative representation of the Brownian local time $\ell(t)$ at 0, also due to Lévy (1948), and proved in Chapter 20. Namely, that $2\ell(t, \cdot)$ has the same distribution as the maximum M_t of a standard Brownian motion $\{B_s : s \geq 0\}$ on $[0, t]$, starting at zero. Also the reflecting Brownian motion X_t has the representation (in law)[8] as

[8] Bhattacharya and Waymire (2021), Theorem 19.3, p. 230.

$M_t - B_t$, and the joint distribution is the same as that of $(M_t, M_t - B_t)$. So that, using the joint distribution[9] of (M_t, B_t), one has

$$\begin{aligned}
g(0) &= \mathbb{E}_0 \int_{[0,\infty)} e^{-\lambda t} e^{-\gamma M_t} f(M_t - B_t)dt \\
&= \int_{[0,\infty)} e^{-\lambda t} dt \int_{[0,\infty)} e^{-\gamma z} dz \\
&\quad \cdot \int_{(-\infty,z]} f(z-x) \frac{2(2z-x)}{\sqrt{2\pi t^3}} e^{-(2z-x)^2/2t} dx \\
&= \int_{[0,\infty)} e^{-\gamma z} \int_{(-\infty,z]} e^{-\sqrt{2\lambda}(2z-x)} f(z-x) dx dz \\
&= (\gamma + \sqrt{2\lambda})^{-1} \int_{[0,\infty)} e^{-\sqrt{2\lambda} z} f(z) dz, \qquad (21.128)
\end{aligned}$$

which is the same as (21.126). For the calculation of the time integral in the third line, see Bhattacharya and Waymire (2021), Corollary 7.12, Proposition 7.15, and Remark 7.8. For the integral in the fourth line, make the transformation $(x, z) \to y = (z - x, z)$ to get

$$\begin{aligned}
&\int_{[0,\infty)} e^{-\gamma z} \int_{(-\infty,z]} e^{-\sqrt{2\lambda}(2z-x)} f(z-x) dx dz \\
&= \int_{[0,\infty)} e^{-2\gamma z} \int_{(0\infty)} e^{-\sqrt{2\lambda} y} e^{-\sqrt{2\lambda} z} f(y) dx dy \\
&= \int_{([0,\infty)} e^{-(\sqrt{2\lambda}+\gamma) z} dz \int_{[0,\infty)} e^{-\sqrt{2\lambda} y} f(y) dy.
\end{aligned}$$

■

Remark 21.7 It is clear from the long proof above that a precise probabilistic representation under the boundary condition (21.113), or (21.127), that for general diffusions such a representation would be difficult to obtain. Feller (1954) shows that in the general case, a diffusion with the elastic boundary condition may be constructed as a limiting process. When r_0 is reached, let the jump to a point x_0 in the interior with probability θ_0 and with probability $1-\theta_0$ the process terminates. With a probability κ_0, this process then vanishes on return to r_0, and with probability $1-\kappa_0$ continues as described before. One lets $x_{\kappa_0} \to r_0$ and $\theta_0 \to 1$ (while κ_0 remains fixed); then the limiting process has boundary condition at r_0 of the type (21.127). Feller provides a detailed and rather long derivation of the constant γ.

[9] See Bhattacharya and Waymire (2021), Corollary 7.12, p. 82.

Remark 21.8 (Time Change) There is a general method, called *time change*, which transforms Brownian motion $\{B_t\}$ into arbitrary diffusions. For a diffusion with inaccessible boundaries in natural scale and speed measure $m(dy)(= dm(y))$, let

$$M(t) := \frac{1}{2}\int_{(-\infty,\infty)} \ell(t, y)m(dy),$$

where $\ell(t, y)$ is the local time of Brownian motion. Let

$$T(t) := \max\{s : M(s) = t\},$$

i.e., T is the inverse of M. Then it may be shown[10] that $\{X_t := B_{T(t)}\}$ is a diffusion in natural scale on $(-\infty, \infty)$. Similar constructions can be made for diffusions with local boundary conditions on half-lines[11] $(0, \infty)$ or $[0, \infty)$.

The problem of changing the original scale function $s(x)$ (see, e.g., (21.4)) to natural scale natural scale and carrying out computations in the natural scale is often a little inconvenient. So we summarize the boundary classification in terms of a given scale $s(x)$ and speed function $m(x)$ here.

(i) A boundary r_i is inaccessible if

$$|\int_{[0,r_i)} m(x)d^+s(x)| = \infty. \tag{21.129}$$

Staring from the interior (r_0, r_1), the probability of reaching the boundary r_i is zero.

(ii) An accessible boundary r_i is regular if

$$|\int_{[0,r_i)} s(x)dm(x)| < \infty. \tag{21.130}$$

(iii) An accessible boundary r_i is an exit boundary if

$$|\int_{[0,r_i)} s(x)dm(x)| = \infty. \tag{21.131}$$

(iv) An inaccessible boundary r_i is natural if, in addition to (21.129), the quantity in (21.131) is also infinite.

(v) An inaccessible boundary r_i is an entrance boundary if (21.131) holds.

The classification in Lemma 6 remains valid when expressed in terms of the scale function $s(x)$ and speed function $m(x)$.

---

[10] See Itô and McKean (1965), pp.167–176.

[11] See Itô and McKean (1963).

Asymptotic behaviors of one-dimensional diffusions, such as (i) recurrence or transience and (ii) positive recurrence or null recurrence, are as analyzed for non-singular diffusions with smooth coefficients in Chapter 11. The proofs given there in terms of scale and speed functions apply to Feller's general diffusions as well. We conclude with a number of examples of one-dimensional diffusions[12]

Example 7 (Brownian Motion)

$$I = [-\infty, \infty], \quad L = \frac{1}{2}\frac{d^2}{dx^2}.$$

Both boundaries are inaccessible and natural. Here $s(x) = x, m(x) = 2x$.

Example 8 (Ornstein–Uhlenbeck Process)

$$I = [-\infty, \infty], \quad L = \beta x \frac{d}{dx} + \frac{1}{2}\sigma^2 \frac{d^2}{dx^2}$$

with β, σ nonzero constants. Both boundaries $-\infty, \infty$ are inaccessible and natural.

Example 9 (Dirichlet, Absorbing, and Reflecting Boundary Conditions)

$$I = [0, \infty], \quad L = \frac{1}{2}\frac{d^2}{dx^2}.$$

Here 0 is a regular boundary and ∞ is a natural boundary. With boundary condition $f(0) = 0, r_0 = 0$ is a regular Dirichlet boundary; under $f'(0^+) = 0$, 0 is a reflecting boundary; and under $f''(0) = 0$, $r_0 = 0$ is an absorbing boundary. In the first case, the diffusion is nonconservative. But in the last two cases, the diffusion is conservative.

Example 10 (Radial Brownian Motion) Let k be an integer, $k \geq 2$, and

$$I = [0, \infty], \quad L = \frac{k-1}{2x}\frac{d}{dx} + \frac{1}{2}\frac{d^2}{dx^2}.$$

Fixing a point, say 1, in the interior of I, one may compute $s(x)$ and $m(x)$ as in (21.4) but with the integrals taken over $[1, x)$. Boundaries are inaccessible, but $r_0 = 0$ is entrance, while $r_1 = \infty$ is natural. This process arises as $X_t = |B_t|, t \geq 0$, where $\{B_t\}$ is a k-dimensional standard Brownian motion (see Chapter 13), $k > 1$. One could include 0 in the state space $(0, \infty)$; but from the interior the probability of reaching 0 is zero, while starting from zero the process would immediately move to the interior, never returning to 0.

[12] See Mandl (1968), Karlin and Taylor (1981) for more examples.

Example 11 (Raleigh Process) This is a more general form of Example 10, with

$$I = [0, \infty], \quad L = \frac{\beta - 1}{2x}\frac{d}{dx} + \frac{1}{2}\frac{d^2}{dx^2}.$$

For $1 \leq \beta < 2$, 0 is a regular boundary, and a boundary condition needs to be attached (e.g., $f(0) = 0$, or $f\prime(0) = 0$, etc.) to define the diffusion; for $\beta \geq 2$, $r_0 = 0$ is an entrance boundary. For $\beta < 1$, 0 is an exit boundary.

Example 12 (Wright–Fisher Gene Frequency Without Mutation)

$$I = [0, 1], \quad L = cx(1-x)\frac{d^2}{dx^2}, \ \ c > 0.$$

Here $r_0 = 0$ and $r_1 = 1$ are accessible and exit boundaries, respectively.

Remark 21.9 The Wright–Fisher diffusion model arises in a scaling limit of a discrete time Markov chain in which x and $1 - x$ are gene frequencies of A-genes and a-genes, respectively, without mutation or selective advantage.

Exercises

1. Let $g, g', \frac{d}{dm}g$ be continuous.
   (a) Show that if $g' = 0$ on (r_0, r_1) or $\frac{d}{dm}g = 0$, then g is a constant on (r_0, r_1). [*Hint*: Let $x < y$, $g(y) - g(x) = \int_{(x,y]} \frac{dg}{dm}(z)dm(z) = 0$.]
   (b) Show that for $x < y$, $g(y) - g(x) = \int_{(x,y]} \left(\int_{(x,z]} \frac{d}{dm}\frac{d}{du}g(u)m(du) \right) dz + (y - x)g'(x)$. [*Hint*: The inner integral equals $g'(z) - g'(x)$.]
   (c) Let g_1, g_2 be two linearly independent solutions of $Lg - \lambda g = 0$. Show that every solution g is a linear combination of g_1, g_2. [*Hint*: There exist unique constants c_1, c_2 such that $c_1 g_1(0) + c_2 g_2(0) = g(0)$, and $c_1 g_1'(0) + c_2 g_2'(0) = g'(0)$; recall that $W(g_1, g_2)$ is a constant. Then consider $W(c_1 g_1 + c_2 g_2, g)$ and show this implies $g = c(c_1 g_1 + c_2 g_2)$ for some constant c.]
2. Suppose that $\{T_t\}$ is a strongly continuous contraction semigroup on $C(I)$.
   (a) Prove that $x \to p(t; x, B)$ is measurable for every $B \in \mathcal{B}(I)$ and $t > 0$.
   (b) Prove that the Laplace transform is one-to-one on the space of all bounded continuous functions on $[0, \infty)$. [*Hint*: Let f be bounded and continuous, and $\hat{f}(\lambda) = \int_{[0,\infty)} e^{-\lambda t} f(t)dt$ its Laplace transform on $(0, \infty)$. Then $f^{(k)}(\lambda) = (-1)^k \int_{[0,\infty)} e^{-\lambda t} t^k f(t)dt$. Consider $\sum_{k=0}^{[\lambda x]} \frac{(-1)^k}{k!} \lambda^k \hat{f}^{(k)}(\lambda) = \int_{[0,\infty)} P(\frac{N_\lambda}{\lambda} \leq \frac{x}{t}) f(t)dt$, where N_λ is Poisson distributed with parameter

λ, so that N_λ/λ converges in distribution to the Dirac measure $\delta_{\{i\}}$. The last integral converges, as $\lambda \to \infty$, to $\int_{[0,x]} f(t)dt$ for all x, and thus determining f. Note that the assumption of continuity may be replaced by measurability with the same conclusion.]

3. Prove that for every $\varepsilon > 0$ and $x \in (r_0, r_1)$, there exists $f_x \in \mathcal{D}$ such that $f_x(y) \geq 0\ \forall y \in I$, and (21.63) holds. [*Hint*: Let $g(y) = 1$ if $|y - x| > \varepsilon/2$, and $g(y) = 0$ if $|y - x| < 1/2$. Now consider the convolution of g with a twice continuously differentiable p.d.f. φ_ε with support in $\{y : |y| \leq \varepsilon/6\}$.]
4. Suppose $\{B_t : t \geq 0\}$ is a standard one-dimensional Brownian motion. It was shown in Chapter 13 that $\{|B_t| : t \geq 0\}$ is a Markov process. Prove that the latter process is generated as in Example 3.
5. (a) Calculate $s(x)$ and $m(x)$ in Example 4, with the point 1, instead of 0, as the pivot.
   (b) Let $\{B_t : t \geq 0\}$ be a k-dimensional standard Brownian motion. The radial Brownian motion defined by $\{|B_t| : t \geq 0\}$ was shown to be a Markov process in Chapter 13. Show that the generator of this process is as in Example 4.
6. Show that the diffusion in Example 6 is that in Example 5 but with 0 absorbing at 1 vanishing.
7. Suppose that $s(\cdot), m(\cdot)$ are given strictly increasing continuous functions on $[0, \infty), s(0) = 0 = m(0)$. A probabilistic method of constructing a diffusion on $[0, \infty)$ with these scale and speed functions, with reflecting boundary condition $f'(0^+) = 0$, and with either inaccessible or local boundary condition at ∞, is the following. First extend the functions $s(\cdot), m(\cdot)$ on $(-\infty, \infty)$ by letting $s(-x) = s(x), m(-x) = -m(x)$ for $x > 0$. Let $\{X_t : t \geq 0\}$ be a diffusion on $(-\infty, \infty)$ with these extended coefficients and the same boundary condition at $-\infty$ as at ∞. Show that the folded process $\{Y_t = |X_t| : t \geq 0\}$ is a diffusion on $[0, \infty)$ with scale and speed functions as prescribed, boundary condition $f'(0) = 0$, and the given boundary condition at ∞, as mentioned.
8. Let $L = \frac{1}{2}\frac{d^2}{dx^2}$ on the open half-line $(0, \infty)$ with absorbing boundary condition $f(0^+) = 0$. Compute $(\lambda - L)^{-1} f(x) = \int_0^\infty G(x, y) f(y) m(dy) + c(f) v_+(x)$, for a suitable constant $c(f)$. [*Hint*: Use Lemma 8(b) to see that $c(f) = 0$.]
9. (a) Verify (21.125). [*Hint*: Use Example 1 to get the Laplace transform of the reflected Brownian motion X_t as $\frac{1}{2\sqrt{2\lambda}}(e^{-\sqrt{2\lambda}|x-y|} + e^{-\sqrt{2\lambda}(y-x)}\mathbf{1}_{[0,y)}(x))$. Then (21.125) is $\int_{(0,\infty)} e^{-\sqrt{2\lambda}y} f(y)dy$.] (b) Check (21.128)
10. (*Skew Brownian Motion*) Fix $0 < \alpha < 1$. The α-skew Brownian motion can be defined as the diffusion on $\mathbb{R}$ having generator $L = \frac{d}{dmd^+s}$ for the scale function

$$s(x) = \begin{cases} 2(1-\alpha)^{-1}x, & \text{if } x < 0 \\ 2\alpha^{-1}x, & \text{if } x \geq 0 \end{cases},$$

and speed function

$$m(x) = \begin{cases} (1-\alpha)x, & \text{if } x < 0 \\ \alpha x, & \text{if } x \geq 0 \end{cases}.$$

Show that

$$\mathcal{D}_L = \{f \in C(\mathbb{R}) : (1-\alpha)f'(0^-) = \alpha f'(0^+),\ f'' \text{ exists \& continuous on} \mathbb{R}\backslash\{0\}\}.$$

11. (*Branching Diffusion*) Compute the scale and speed functions, and verify the boundary conditions for the diffusion having generator of the form $I = [0, 1]$, $L = \beta \frac{d^2}{dx^2} + \alpha x \frac{d}{dx}$, where α, β are positive constants.
12. Assume natural scale, but suppose that the speed measure has a positive mass at x_0. For $f \in \mathcal{D}_L$, show that $f'(x_0^+) - f'(x_0^-) = \int_I h(y)m(dy)$ for some $h \in C(I)$. [*Hint*: $f \in \mathcal{D}_L$ is in the range of the resolvent.]
13. Assume $s(x) = x$, $x \in (-\infty, \infty)$, with inaccessible boundaries $\pm\infty$. Assume $G(x, y) = G(-x, y)$ for all x, y.
    (i) Show $G(x, y) = G(-x, -y) = G(x, -y)$ for all x, y. [*Hint*: Use symmetry of $G(x, y)$.]
    (ii) Show that $p(t; x, dy) = p(t; -x, dy)$, $x \geq 0$.
14. (*Slowly Reflecting (or Sticky) Point at* 0) Let $s(x) = x$, $m(dx) = 2dx + 2\gamma\delta_{\{0\}}(dx)$, $-\infty < x < \infty$, for a given parameter $\gamma > 0$. The speed function with $m(0) = 0$ is

$$m(x) = \begin{cases} 2x & \text{if } x \geq 0, \\ 2x - 2\gamma & \text{if } x < 0 \end{cases}.$$

Fix an arbitrary resolvent parameter $\lambda > 0$.

(i) Show

$$g(x; \lambda) = \begin{cases} \cosh(\sqrt{2\lambda}x) & x \geq 0 \\ \cosh(\sqrt{2\lambda}x) - \sqrt{2\lambda}\gamma \sinh(\sqrt{2\lambda}x) & x < 0. \end{cases}$$

[*Hint*: Make the recursive computation in (21.18).]

(ii) Show

$$g_+(x) = \begin{cases} e^{-\sqrt{2\lambda}x} & \text{if } x \geq 0 \\ e^{-\sqrt{2\lambda}x} - \gamma\sqrt{2\lambda}\sinh(\sqrt{2\lambda}x) & \text{if } x < 0, \end{cases}$$

and

$$g_-(x) = \begin{cases} e^{\sqrt{2\lambda}x} + \gamma\sqrt{2\lambda}\sinh(\sqrt{2\lambda}x) & \text{if } x \geq 0, \\ e^{\sqrt{2\lambda}x} & \text{if } x < 0. \end{cases},$$

$$W(g_-, g_+) = 2(\lambda\gamma + \sqrt{2\lambda}).$$

[*Hint*: The integrals involving hyperbolic trigonometric functions are computable in closed form by simple substitution methods, and $G(x, y)$ is unchanged by scaling g_+ or g_- by nonzero constants.]

(iii) Show that

$$G(x, y) = \begin{cases} \frac{2+\gamma\sqrt{2\lambda}}{4(\sqrt{2\lambda}+\gamma\lambda)} e^{-\sqrt{2\lambda}|y-x|} - \frac{\gamma\sqrt{2\lambda}}{4(\sqrt{2\lambda}+\gamma\lambda)} e^{-\sqrt{2\lambda}|x+y|}, x \cdot y \geq 0 \\ \frac{1}{2(\sqrt{2\lambda}+\gamma\lambda)} e^{-\sqrt{2\lambda}|y-x|}, x \cdot y < 0. \end{cases}$$

(iv) Show that $G(x, y)$ is unchanged in the case of the natural scale function $s(x) = x$, and odd speed function

$$m(x) = \begin{cases} 2x + \gamma & \text{if } x \geq 0, \\ 2x - \gamma & \text{if } x < 0 \end{cases}.$$

[*Hint*: Adding a constant to the speed function does not change the speed measure.]

(v) Show[13] that the set $\{t \geq 0 : X_t = 0\}$ of zeroes of sticky Brownian motion has positive Lebesgue measure. [*Hint*: Consider (21.58) and (21.59), and check that $\mathbb{E}_0 \int_0^\infty e^{-\lambda t}\mathbf{1}_{[X_t \in \mathbb{R}\setminus\{0\}]}dt < \frac{1}{\lambda}$ by an exact calculation of $\int_{\mathbb{R}} G(0, y)\mathbf{1}_{\mathbb{R}\setminus\{0\}}(y)m(dy) = \frac{\sqrt{2\lambda}}{\sqrt{2\lambda}+\gamma\lambda}\frac{1}{\lambda}$. Note that $\int_{\mathbb{R}} G(0, y)m(dy) = \frac{1}{\lambda}$.]

15. Consider Feller's diffusion X on the half-line $I = [0, \infty)$ with natural scale and speed functions $s(x) = x, m(x) = 2x, x > 0$, i.e., $L = \frac{1}{2}\frac{d^2}{dx^2}$ on $(0, \infty)$, respectively, with boundary condition $f'(0+) = \gamma f''(0+)$, for a positive nonnegative constant γ. This is often viewed as an interpolation between reflecting and absorbing boundary conditions. For $\gamma = 0$ one obtains a reflecting boundary at $x = 0$, i.e., $f'(0^+) = 0$, and absorption may be viewed as the case $\gamma = \infty$, i.e., $f''(0^+) = \frac{1}{\gamma}f'(0^+) = 0$. Show that the zero set $\{t \geq 0 : X_t = 0\}$ has positive Lebesgue measure for $0 < \gamma < \infty$. [*Hint*: Note

[13] Slowly reflecting behavior was discovered by Feller (1952). See Peskir (2015) for a historical account. See Salminen and Stenlund (2021) for this and other interesting examples. Also see Bass (2014) for an analysis of the corresponding stochastic differential equation.

that $(\lambda - L)^{-1} f(x) = Gf(x) + cg_+(x)$, where $c = -\frac{-(Gh)'(0)+\gamma(Gh)''(0)}{-g'_+(0)+\gamma g''_+(0)}$. From here apply the same hint as in Exercise 14.]

16. Let $I = [0, \infty)$, and $L = \frac{1}{2}\frac{d^2}{dx^2}$ on $C^2(I)$ together with each case of Feller's general boundary condition, $\alpha f(0) - \gamma f'(0) + \beta f''(0) = 0$, $\alpha, \beta, \gamma \geq 0$, $\alpha + \beta + \gamma = 1$, enumerated below. In each case, describe the respective sample path behavior of the restricted Brownian motion upon reaching the boundary $r_0 = 0$. (i) $f(0) = 0$. (ii) $f''(0) = 0$. (iii) $f'(0) = 0$. (iv) $\alpha f(0) - \gamma f'(0) = 0$, $\quad (\alpha\beta > 0)$. (v) $\beta f''(0) = \gamma f'(0)$, $\quad (\beta\gamma > 0)$. [*Hint*: Review the chapter examples and Exercise 15.]
17. Check that the proofs of (i) transience or recurrence and (ii) null recurrence or positive recurrence, given in Chapter 8, apply to Feller's general one-dimensional diffusions.
18. (a) Let $\mathscr{X}$ be a Banach space and f an unbounded linear functional on $\mathscr{X}$. Prove that the kernel of f is dense in $\mathscr{X}$. [*Hint*: Suppose $x \notin \ker(f)$. There is a sequence $\{x_n\}$ such that $|x_n| = 1$, $x_n > n$ for all $n = 1, 2, \ldots$. Consider $y_n = x - \frac{f(x)}{f(x_n)} x_n$. Then $y_n \in \ker(f)$ and $y_n \to x$.] (b) Prove that the linear functionals $f'(r_i)$ and $Lf(r_i)$ are unbounded in $C(I)$. (c) In the case of a Dirichlet condition $f(r_i) = 0$, the semigroup is strongly continuous on a proper closed subspace $C_0(I)$ of $C(I)$.

Chapter 22
Eigenfunction Expansions of Transition Probabilities for One-Dimensional Diffusions

We consider one-dimensional non-singular diffusions with Dirichlet or Neumann (reflecting) boundary conditions. The generators of these diffusions are self-adjoint on $L^2([r_0, r_1], m(dx))$, where $m(dx)$ is the speed measure on the state space $I = [r_0, r_1]$. This leads to eigenfunction expansions of the transition probability densities of the diffusions with these generators.

Let $L = \frac{d^2}{dmd^+s}$ be the infinitesimal generator of a Feller diffusion on $C(I)$, $I = [r_0, r_1]$, with Dirichlet or Neumann (reflecting) boundary conditions on $r_i (i = 0, 1)$. Here, $s(x)$ is the scale function, and $m(x)$ is the speed function, as defined in Chapter 21, and $m(dx)$ is the speed measure with density $m'(x)$. Recall that r_0, r_1 are finite in this case, as are $s(x)$ and $m(x)$. Feller's results apply under more general regular boundary conditions, but, for simplicity, we only consider the Dirichlet and reflecting boundary conditions. We will denote by $\mathcal{D}_L$ the domain of the generator of the corresponding diffusion on $L^2([r_0, r_1], m(dx))$, extending it from the generator on $C(I)$ considered in Chapter 21. Let $p(t; x, dy)$ denote the transition probability, or sub-probability (depending on the boundary condition). This also defines the semigroup $\{T_t : t \geq 0\}$, on $L^2([r_0, r_1], m(dx))$:

$$T_t f(x) = \int_{[r_0,r_1]} f(y)p(t; x, dy), \quad f \in L^2([r_0, r_1], m(dx)). \tag{22.1}$$

It is a strongly continuous contraction semigroup on the Hilbert space $L^2[r_0, r_1] := L^2([r_0, r_1], m(dx))$, or on a subspace $L_0^2[r_0, r_1]$ if at least one of the boundaries has the Dirichlet boundary condition. Strong continuity holds, since on the Banach space $C(I)$, the sup norm $||(\lambda - L)^{-1}||$ is larger (i.e., no smaller) than the L^2-norm

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_22

$||(\lambda - L)^{-1}||_2$ restricted to $C(I)$, and $||(\lambda - L)^{-1}|| \leq \frac{1}{\lambda}$, as proved in Chapter 21. It follows from the Hille-Yosida Theorem that $\{T_t : t \geq 0\}$ is a contraction semigroup in $L^2[r_0, r_1]$; also see Exercise 1 for another proof of contraction.

For simplicity, we will make the following:

$$\textbf{Assumption} : s(x) \text{ and } m(x) \text{ are continuously differentiable.} \tag{22.2}$$

Theorem 22.1 *Under Dirichlet or Neumann boundary conditions, the semigroup* $\{T_t : t \geq 0\}$ *and its generator are self-adjoint in* $L^2[r_0, r_1]$*, or a proper subspace of it in case at least one of the boundaries has Dirichlet condition.*

Proof Let $f, g \in \mathcal{D}_L$, under Dirichlet or Neumann boundary conditions on r_i (i.e., $f(r_i) = 0$ or $f'(r_i) = 0$, and same for g), $i = 0, 1$. Then, writing the inner product in $L^2[r_0, r_1]$, as $\langle f, g\rangle$, one has

$$\begin{aligned}
\langle Lf, g\rangle &= \int_{[r_0,r_1]} (\frac{d^2}{dmds} f)(x)g(x)m(dx) \\
&= \int_{[r_0,r_1]} \frac{d}{dm}(\frac{1}{s'(x)} f'(x))g(x)m(dx) \\
&= \int_{[r_0,r_1]} \frac{1}{m'(x)} \frac{d}{dx}[\frac{1}{s'(x)} f'(x)]g(x)m'(x)dx \\
&= \int_{[r_0,r_1]} \frac{d}{dx}[\frac{1}{s'(x)} f'(x)]g(x)dx \\
&= \int_{[r_0,r_1]} \frac{1}{s'(x)} f'(x)g'(x)dx,
\end{aligned} \tag{22.3}$$

using the boundary conditions $f'(r_i)g(r_i) = 0$ for $i = 0, 1$. The last expression in (22.3) is symmetric in f and g. ■

The next result provides for eigenfunction expansions of transition probability densities under the Dirichlet or Neumann boundary conditions.

Theorem 22.2 *Under Dirichlet or Neumann boundary conditions on* $r_i (i = 0, 1)$*: (a) For all* $\lambda > 0$*, the resolvent operator* $f \to (\lambda - L)^{-1} f = \int_{[0,\infty)} e^{-\lambda t} T_t f dt$ *is compact on* $L^2[r_0, r_1]$ *or on a proper subspace of it in the case of Dirichlet boundary conditions. (b) The transition probability, or sub-probability, density* $p(t; x, y)$ *with respect to* $m(dy)$ *has an absolutely convergent expansion* $\sum_{n\geq 0} \exp\{\alpha_n t\}\varphi_n(x)\varphi_n(y)$*, where* α_n *are the nonpositive eigenvalues of* L *with corresponding normalized eigenfunctions* φ_n*. The convergence is uniform on* $I \times I$*.*

Proof The self-adjoint Green's operator $f \to \widetilde{G} f$ with kernel $\sum_{n\geq 0} \widetilde{G}(x, y)$, say, under the boundary conditions imposed, is smaller than $G(x, y)$, as argued in the proof of Theorem 21.6. Therefore, by a standard theorem in functional analysis (see Appendix A in Bhattacharya and Waymire (2022)), $\widetilde{G}$ is a compact self-adjoint

operator on $H = L^2[r_0, r_1]$, or a proper closed subspace H_0 of it in case of Dirichlet boundary condition. Note also that $\widetilde{G}(x, y) = \int_{[0,\infty)} e^{-\lambda t} p(t; x, y)dt$, and it has an expansion

$$\sum_{n \geq 0} \beta_n \varphi_n(x)\varphi_n(y),$$

where β_n are the nonnegative eigenvalues of the Green's operator $\widetilde{G}$ with corresponding normalized eigenfunctions φ_n $(n \geq 0)$, $\sum \beta_n^2 < \infty$ (Theorem A.3 in Appendix A, Bhattacharya and Waymire (2022)). The orthonormal sequence is complete in the Hilbert space, because the domain of $\widetilde{G}$ is dense in the Hilbert space. Because $\widetilde{G} = (\lambda - L)^{-1}$, and for $u \in \mathcal{D}_L$, $u = (\lambda - L)^{-1} f$, one has

$$u(x) = \sum_n \int_I \beta_n \varphi_n(x)\varphi_n(y) f(y)m(dy),$$

and

$$Lu(x) = \lambda u(x) - f(x) = \sum_n (\lambda\beta_n - 1)\langle f, \varphi_n\rangle \varphi_n(x).$$

In particular,

$$L\varphi_n(x) = (\lambda\beta_n - 1)\varphi_n(x).$$

That is, $\alpha_n := \lambda\beta_n - 1$ are eigenvalues of L with corresponding eigenfunctions φ_n $(n = 0, 1, \ldots)$. Next consider the equation

$$\frac{d}{dt} T_t f = LT_t f = T_t Lf \quad (f \in \mathcal{D}_L).$$

Specializing to $f = \varphi_n$, one obtains $\frac{d}{dt} T_t\varphi_n = \alpha_n T_t\varphi_n$, yielding $T_t\varphi_n = \exp\{t\alpha_n\}\varphi_n(\forall n)$. Hence

$$T_t f = \sum_{n \geq 0} \exp\{\alpha_n t\}\varphi_n(x) \int_I \varphi_n(y) f(y)m(dy).$$

This holds for all f in the Hilbert space H or H_0, as the case may be. It follows that the density $p(t; x, y)$ (with respect to $m(dy)$) has the expansion

$$p(t; x, y) = \sum_{n \geq 0} \exp\{\alpha_n t\}\varphi_n(x)\varphi_n(y), \tag{22.4}$$

the convergence being absolute and uniform. ■

Example 1 (Brownian Motion with Two Reflecting Boundaries) In this case, $I = [0, d]$, $x_0 = 0$, $s(x) = x$, $m(x) = 2x/\sigma^2$, and

$$L = \frac{\sigma^2}{2}\frac{d^2}{dx^2}, \tag{22.5}$$

with the *backward boundary condition*

$$\left.\frac{d}{dx}\right|_{x=0,d} = 0 \qquad (t > 0;\ y \in I). \tag{22.6}$$

We seek eigenvalues α of L and corresponding eigenfunctions φ. That is, consider solutions of

$$\frac{1}{2}\sigma^2\varphi''(x) = \alpha\varphi(x), \tag{22.7}$$

$$\varphi'(0) = \varphi'(d) = 0. \tag{22.8}$$

Check that the functions

$$\varphi_n(x) = b_m \cos(n\pi\,\frac{x}{d}) \qquad (n = 0, 1, 2, \dots), \tag{22.9}$$

satisfy (22.7), (22.8), with α given by

$$\alpha = \alpha_n = -\frac{n^2\pi^2\sigma^2}{2d^2} \qquad (n = 0, 1, 2, \dots). \tag{22.10}$$

Now $m'(x) = 2/\sigma^2$. Since this is a constant, one may use dy in place of dm in Theorems 22.1, 22.2. Using

$$\int_0^d 1^2 dx = d, \qquad \int_0^d \cos^2(\frac{n\pi x}{d})dx = \frac{d}{2}, \tag{22.11}$$

the normalizing constant b_n is given by

$$b_n = \sqrt{\frac{2}{d}} \quad (n \geq 1); \quad b_0 = \sqrt{\frac{1}{d}}. \tag{22.12}$$

To explicitly see that (22.9) and (22.10) provide all the eigenvalues and eigenfunctions, one can check the *completeness* of $\{\varphi_n : n = 0, 1, 2, \dots\}$ in $L^2([0, d], dy)$. This may be done by using Fourier series (Exercise 9). Thus, for $t > 0$, $0 < x, y < d$,

$$p(t; x, y) = \sum_{n=0}^{\infty} e^{\alpha_n t} \varphi_n(x)\varphi_n(y)$$
$$= \frac{1}{d} + \frac{2}{d} \sum_{n=1}^{\infty} \exp\{-\frac{\sigma^2 \pi^2 n^2}{2d^2} t\} \cos(\frac{n\pi x}{d}) \cos(\frac{n\pi y}{d}). \tag{22.13}$$

Notice that

$$\lim_{t\to\infty} p(t; x, y) = \frac{1}{d}, \tag{22.14}$$

the convergence being *exponentially fast*, uniformly with respect to x, y. Thus, as $t \to \infty$, the distribution of the process at time t converges to the uniform distribution on $[0, d]$. In particular, this limiting distribution provides the invariant initial distribution for the process.

Example 2 (Brownian Motion with Two Absorbing Boundaries) Here $I = [0, d]$, $x_0 = 0$, $s(x) = \frac{\sigma^2}{2\mu}(1 - e^{-\frac{2\mu}{\sigma^2}x})$, $m(x) = \frac{\sigma^2}{\mu^2}(e^{\frac{2\mu}{\sigma^2}x} - 1)$, and the transition probability has a density $p(t; x, y)$ on $0 < x, y < d$.

$$L = \frac{\sigma^2}{2} \frac{d^2}{dx^2} + \mu \frac{d}{dx}, \tag{22.15}$$

with backward boundary conditions on $u \in \mathcal{D}_L$,

$$u(0) = 0 \quad u(d) = 0. \tag{22.16}$$

for $t > 0$. Check that the functions

$$\varphi_n(x) = b_n \exp\{-\frac{\mu x}{\sigma^2}\} \sin(\frac{n\pi x}{d}) \qquad (n = 1, 2, \dots) \tag{22.17}$$

are eigenfunctions corresponding to eigenvalues

$$\alpha = \alpha_n = -\frac{n^2\pi^2\sigma^2}{2d^2} - \frac{\mu^2}{2\sigma^2}, \qquad (n = 1, 2, \dots). \tag{22.18}$$

Since

$$m'(x) = \exp\{\frac{2\mu x}{\sigma^2}\} \qquad (0 \le x \le d), \tag{22.19}$$

the normalizing constants are

$$b_n = \sqrt{\frac{2}{d}} \qquad (n = 1, 2, \dots). \tag{22.20}$$

The eigenfunction expansion of the density function $p(t; x, y)$ with respect to $m(dy) = m'(y)dy$ is therefore

$$\begin{aligned} p(t; x, y) &= \sum_{n=1}^{\infty} e^{\alpha_n t} \varphi_n(x)\varphi_n(y)\pi(y) \\ &= \frac{2}{d} \exp\{\frac{-\mu(y+x)}{\sigma^2}\} \exp\{-\frac{\mu^2 t}{2\sigma^2}\} \sum_{n=1}^{\infty} \exp\{-\frac{n^2\pi^2\sigma^2 t}{2d^2}\} \\ &\quad \times \sin(\frac{n\pi x}{d}) \sin(\frac{n\pi y}{d}) \qquad (t > 0,\ 0 < x, y < d). \end{aligned} \tag{22.21}$$

The density $\tilde{p}(t; x, y)$, say, with respect to Lebesgue measure dy is

$$\begin{aligned} \tilde{p}(t; x, y) &= p(t; x, y)m'(y) \\ &= p(t; x, y) 2 \exp\{\frac{2\mu y}{\sigma^2}\} \\ &= \frac{2}{d} \exp\{\mu(y-x)/\sigma^2\} \exp\{-\frac{\mu^2}{2\sigma^2}\} \exp\{-\frac{\mu}{2\sigma^2} t\} \\ &\quad \times \sum_{n=1}^{\infty} \exp\{-\frac{n^2\pi^2\sigma^2}{2d^2} t\} \sin(\frac{2\pi x}{d}) \cos(\frac{2\pi y}{d}), \end{aligned} \tag{22.22}$$

for $t > 0, 0 < x, y < d$. It remains to calculate

$$p(t; x, \{0\}) = P_x(X_t = 0), \quad p(t; x, \{d\}) = P_x(X_t = d), \quad 0 < x < d. \tag{22.23}$$

Note that

$$\begin{aligned} p(t; x, \{0\}) + p(t; x, \{d\}) &= 1 - \int_{(0,d)} \tilde{p}(t; x, y) dy \\ &= 1 - \frac{2}{d} \exp\{-\frac{\mu x}{\sigma^2}\} \\ &\quad \times \sum_{n=1}^{\infty} a_n \exp\{-\frac{n^2\pi^2\sigma^2 t}{2d^2}\} \sin(\frac{n\pi x}{d}) \end{aligned} \tag{22.24}$$

where

$$a_n = \int_0^d \exp\{\frac{\mu y}{\sigma^2}\}\cos(\frac{n\pi y}{d})dy. \tag{22.25}$$

Thus, it is enough to determine $p(t; x, \{0\})$; the function $p(t; x, \{d\})$ is then determined from (22.24). Thus,

$$p(t; x, \{0\}) = \varphi(x) - \int_{(0,d)} \varphi(y)\tilde{p}(t; x, y)dy \tag{22.26}$$

where $\varphi(x) = P_x(\{X_t\}$ reaches 0 before $d)$. In particular,

$$\varphi(x) = \begin{cases} \frac{d-x}{d} \text{ if } \mu = 0 \\ \frac{1-\exp\{2\mu(d-x)/\sigma^2\}}{1-\exp\{2d\mu/\sigma^2\}} \text{ if } \mu \neq 0. \end{cases} \tag{22.27}$$

Substituting (22.21), (22.27) in (22.26) yields $p(t; x, \{0\})$.

As further application of the eigenfunction expansions on compact intervals, one may obtain formulae on the half-line by passage to a limit as illustrated by the examples to follow.

Example 3 (Brownian Motion on Half-line with One Reflecting Boundary) Let $I = [0, \infty)$, $x_0 = 0$, $s(x) = x$, $m(x) = \frac{2}{\sigma^2}x$. We seek the transition density p with respect to Lebesgue measure, since $m'(y)$ is a constant. Then one has

$$L = \frac{\sigma^2}{2}\frac{d^2}{dx^2}, \quad I = [0, \infty), \quad \frac{d}{dx}|_{x=0} = 0. \tag{22.28}$$

This diffusion is the limiting form of that in Example 1 as $d \uparrow \infty$. Letting $d \uparrow \infty$ in (22.13) one has

$$\begin{aligned} p(t; x, y) &= \lim_{d\uparrow\infty} \frac{1}{d} \sum_{n=-\infty}^{\infty} \exp\{-\frac{\sigma^2\pi^2(n/d)^2}{2}t\}\cos(\frac{n\pi x}{d})\cos(\frac{n\pi y}{d}) \\ &= \int_{-\infty}^{\infty} \exp\{-\frac{t\sigma^2\pi^2u^2}{2}\}\cos(\pi xu)\cos(\pi yu)du \\ &= \frac{1}{2}\int_{-\infty}^{\infty} \exp\{-\frac{t\sigma^2\pi^2u^2}{2}\}[\cos(\pi(x+y)u) + \cos(\pi(x-y)u)]du \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} (e^{-i\xi(x+y)} + e^{-i\xi(x-y)})\exp\{-\frac{t\sigma^2\xi^2}{2}\}d\xi, \qquad (\xi = \pi u) \\ &= \frac{1}{(2\pi t\sigma^2)^{1/2}}(\exp\{-\frac{(x+y)^2}{2t\sigma^2}\} + \exp\{-\frac{(x-y)^2}{2t\sigma^2}\}) \\ &\qquad (t > 0,\ 0 \le x, y < \infty). \end{aligned} \tag{22.29}$$

The last equality in (22.29) follows from Fourier inversion and the fact that $\exp\{-t\sigma^2\xi^2/2\}$ is the characteristic function of the normal distribution with mean zero and variance $t\sigma^2$.

Example 4 (Brownian Motion on Half-line with Absorbing Boundary) $I = [0, \infty)$, $x_0 = 0$, $s(x) = \frac{\sigma^2}{2\mu}(1 - e^{-\frac{2\mu}{\sigma^2}x})$, $m(x) = \frac{\sigma^2}{\mu^2}(e^{\frac{2\mu}{\sigma^2}x} - 1)$. The transition density function $\tilde{p}(t; x, y)$ with respect to Lebesgue measure for $x, y \in (0, \infty)$ is obtained as a limit of (22.21) as $d \uparrow \infty$. That is,

$$\begin{aligned}
p(t; x, y) &= 2\exp\{\frac{\mu(y-x)}{\sigma^2} - \frac{\mu^2 t}{2\sigma^2}\}\int_0^\infty \exp\{-\frac{\pi^2\sigma^2 t u^2}{2}\}\sin(\pi x u)\sin(\pi y u)du \\
&= \frac{2}{\pi}\exp\{\frac{\mu(y-x)}{\sigma^2} - \frac{\mu^2 t}{2\sigma^2}\}\int_0^\infty \exp\{-\frac{\sigma^2 t\xi^2}{2}\}\sin(\xi x)\sin(\xi y)d\xi \\
&= \frac{1}{\pi}\exp\{\frac{\mu(y-x)}{\sigma^2} - \frac{\mu^2 t}{2\sigma^2}\}\int_0^\infty \exp\{-\frac{\sigma^2 t\xi^2}{2}\} \\
&\quad \times[\cos(\xi(x-y)) - \cos(\xi(x+y))]d\xi \\
&= \frac{1}{2\pi}\exp\{\frac{\mu(y-x)}{\sigma^2} - \frac{\mu^2 t}{2\sigma^2}\}\int_{-\infty}^\infty (e^{-i\xi(x-y)} - e^{-i\xi(x+y)})e^{-\sigma^2 t\xi^2/2}d\xi \\
&= \exp\{\frac{\mu(y-x)}{\sigma^2} - \frac{\mu^2 t}{2\sigma^2}\}\frac{1}{(2\pi\sigma^2 t)^{1/2}} \\
&\quad \times[\exp\{-\frac{(x-y)^2}{2\sigma^2 t}\} - \exp\{-\frac{(x+y)^2}{2\sigma^2 t}\}]. \qquad (22.30)
\end{aligned}$$

Integration of (22.30) with respect to y over $(0, \infty)$ yields $p(t; x, \{0\}^c)$, which is the probability that, starting from x, the first passage time to zero is greater than t.

Note that Examples 2 and 4 yield the distributions of the maximum, and the minimum and the joint distribution of the maximum, minimum, and the state at time t of a diffusion with constant coefficients over the time period $[0, t]$ (Exercises 4–5).

Exercises

1. (a) Prove that the semigroup $\{T_t : t \geq 0\}$ under reflecting boundary conditions at r_0, r_1, is contracting in $L^2([r_0, r_1], dm)$. [*Hint*: $||T_t f||_2^2 = \int_I(\int_I f(y)p(t; x, dy))^2 m(dx) = \int_I\int_I f^2(y)p(t; x, dy)m(dx) = \int_I f^2(y)m(dy)$.]

(b) Prove that if at least one of r_i has the Dirichlet boundary condition, while the other has a Dirichlet or Neumann boundary condition, then $\{T_t : t \geq 0\}$ is a contracting semigroup on the closed subspace $H_0 \subset L^2[r_0, r_1]$. [*Hint*: The transition probability is smaller than that in part (a).]

2. Let $p(t; x, dy)$ be the transition probability of a Markov process on a state space $(S, \mathcal{S})$ which is strongly continuous and contracting on an appropriate real Banach space $\mathscr{X}$. Prove that the spectrum of the semigroup is contained in $(-\infty, 0]$. [*Hint*: Every $\lambda > 0$ belongs to the resolvent set.]
3. Assume that the nonzero eigenvalues of L are bounded away from zero.

   (i) In the case of two reflecting boundaries, prove that

   $$\lim_{t\to\infty} p(t; x, y) = \pi(y) \qquad (x, y \in [0, d])$$

   where $\pi(y) = m'(y)/m(I)$. Show that this convergence is exponentially fast, uniformly in x, y.

   (ii) Prove that $\pi(y)dy$ is the unique invariant probability for p.

   (iii) Prove (ii) without the assumption concerning eigenvalues.

4. Use Example 4 to find

   (i) The distribution of the minimum m_t of a Brownian motion $\{X_s\}$ over the time interval $[0, t]$

   (ii) The distribution of the maximum M_t of $\{X_t\}$ over $[0, t]$

   (iii) The joint distribution of (X_t, m_t)

5. Use Example 2 to compute the joint distribution of (X_t, m_t, M_t), using the notation of Exercise 4 above.
6. For a Brownian motion with drift μ and diffusion coefficient $\sigma^2 > 0$, compute the p.d.f of $\tau_a \wedge \tau_b$ when $X_0 = x$ and $a < x < b$.
7. Establish the identities

$$(2\pi\sigma^2 t)^{-1/2} \sum_{n=-\infty}^{\infty} [\exp\{-\frac{(2nd + y - x)^2}{2\sigma^2 t}\} + \exp\{-\frac{(2nd + y + x)^2}{2\sigma^2 t}\}]$$

$$= \frac{1}{d} + \frac{2}{d} \sum_{n=1}^{\infty} \exp\{-\frac{\sigma^2\pi^2 n^2 t}{2d^2}\} \cos(\frac{n\pi x}{d}) \cos(\frac{n\pi y}{d}),$$

$$(2\pi\sigma^2 t)^{-1/2} \sum_{n=-\infty}^{\infty} [\exp\{-\frac{(2nd + y - x)^2}{2\sigma^2 t}\} - \exp\{-\frac{(2nd + y + x)^2}{2\sigma^2 t}\}]$$

$$= \frac{2}{d} \sum_{n=1}^{\infty} \exp\{-\frac{\sigma^2\pi^2 n^2 t}{2d^2}\} \sin(\frac{n\pi x}{d}) \sin(\frac{n\pi y}{d}), \ (0 \leq x, y \leq d).$$

Use these to derive *Jacobi's identity for the theta function*:

$$\theta(z) := \sum_{n=-\infty}^{\infty} \exp\{-\pi n^2 z\} + \frac{1}{\sqrt{z}} \sum_{n=-\infty}^{\infty} \exp\{-\pi n^2/z\} \equiv \frac{1}{\sqrt{z}}\,\theta(\frac{1}{z}),\ z > 0).$$

[*Hint*: Compare (22.13) and (13.49).]

8. (*Hermite Polynomials and the Ornstein-Uhlenbeck Process*) Consider the generator $L = \frac{1}{2}\,d^2/dx^2 - x\,d/dx$ on the state space $\mathbb{R}^1$.
   (i) Prove that L is symmetric on an appropriate dense subspace of $L^2(\mathbb{R}^1, e^{-x^2}dx)$.
   (ii) Check that L has eigenfunctions

   $$H_n(x) := (-1)^n \exp\{x^2\}(d^n/dx^n)\exp\{-x^2\},$$

   the so-called Hermite polynomials, with corresponding eigenvalues $n = 0, -1, -2, \ldots$.
   (iii) Give some justification for the expansion

   $$p(t; x, y) = e^{-y^2} \sum_{n=0}^{\infty} c_n e^{-nt} H_n(x) H_n(y), \qquad c_n := \frac{1}{\sqrt{\pi}2^n n!}.$$

9. According to the theory of Fourier series, the functions $\cos nx$ ($n = 0, 1, 2, \ldots$) and $\sin nx$ ($n = 1, 2, \ldots$) form a complete orthogonal system in $L^2([-\pi, \pi], dx)$. Use this to prove the following:
   (i) The functions $\cos nx$ ($n = 0, 1, 2, \ldots$) form a complete orthogonal sequence in $L^2([0, \pi], dx)$. [*Hint*: Let $f \in L^2([0, \pi], dx)$. Make an even extension of f to $[-\pi, \pi]$, and show that this f may be expanded in $L^2([-\pi, \pi], dx)$ in terms of $\cos nx$ ($n = 0, 1, \ldots$).]
   (ii) The functions $\sin x$ ($n = 1, 2, \ldots$) form a complete orthogonal sequence in $L^2([0, \pi], dx)$. [*Hint*: Extend f to $[-\pi, \pi]$ by setting $f(-x) = -f(x)$ for $x \in [-\pi, 0)$.]

Chapter 23
Special Topic: The Martingale Problem

The martingale problem as formulated by Stroock and Varadhan provides an alternative approach to semigroup theory when defining a diffusion with a given infinitesimal generator by requiring a large class of functions of the process satisfy a natural martingale condition related to the infinitesimal operator for Markov processes. The basic elements of this theory is presented with an application to the approximation of a one-dimensional diffusion by a scaled birth-and-death Markov chain.

To appreciate a basic role of martingales in the theory of Markov processes, consider a *Markov semigroup*, i.e., a strongly continuous contraction semigroup $\{T_t : t \geq 0\}$ on the Banach space $C_b(S)$, where S is Polish. The following result has been referred to as *Dynkin's martingale* in Chapter 17.

Theorem 23.1 *If* $f \in D_A$*, the domain of the generator* A *of the Markov processes* $\{X(t) : t \geq 0\}$*, then* $M(t) = f(X(t)) - \int_{[0,t]} Af(X(u))du$, $t \geq 0$*, is a martingale*

Remark 23.1 One may note by reversing the proof that the martingale relation is equivalent to the semigroup relation $T_t f - f = \int_{(0,t]} T_s Af \, ds$ (for all f in D_A).

In his classic work, Feller (1954) constructed the most general one-dimensional (regular) Markov processes with continuous sample paths by means of semigroup theory and also found the most general boundary conditions for them (see Chapter 21). For multidimensional diffusions, however, semigroup theory for constructing diffusions is not as convenient as that by Itô's seminal theory of stochastic differential equations.

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_23

In this chapter we provide a brief introduction[1] to the so-called martingale problem for diffusions due to Stroock and Varadhan ((1969), (1979)), extending Itô's theory to possibly nonsmooth coefficients. Here, informally, the generator is a second-order elliptic operator:

$$A = \sum_{1\le i\le k} \mu_i(\mathbf{x})\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{1\le i,j\le k} a_{ij}(\mathbf{x})\frac{\partial^2}{\partial x_i \partial x_j}; \qquad [\mathbf{a}(\mathbf{x}) = ((a_{ij}(\mathbf{x})))] \tag{23.1}$$

$\mathbf{a}(\mathbf{x})$ being a nonnegative definite matrix for each $\mathbf{x} = (x_1, \ldots, x_k) \in S = \mathbb{R}^k$. In Itô's seminal theory of stochastic differential equations (SDE) presented in Chapters 6–8, the Markov process governed by A is constructed as the solution of the SDE, expressed equivalently in terms of $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$, $\boldsymbol{\sigma}(\cdot)\boldsymbol{\sigma}'(\cdot) = \mathbf{a}(\cdot)$, $\boldsymbol{\sigma}(\cdot) = ((\sigma_{ij}(\cdot)))_{1\le i,j\le k}$ by

$$d\mathbf{X}(t) = \boldsymbol{\mu}(\mathbf{X}(t))dt + \boldsymbol{\sigma}(\mathbf{X}(t))d\mathbf{B}(t), \tag{23.2}$$

$$\mathbf{X}(t) = \mathbf{X}(0) + \int_{[0,t]} \boldsymbol{\mu}(\mathbf{X}(s))ds + \int_{[0,t]} \boldsymbol{\sigma}(\mathbf{X}(s))d\mathbf{B}(s), \tag{23.3}$$

or,

$$dX_i(t) = \mu_i(\mathbf{X}(t))dt + \sum_{1\le j\le k} \sigma_{ij}(\mathbf{X}(t))dB_j(t)\ (1 \le i \le k). \tag{23.4}$$

Here $\mathbf{B} = (B_1, \ldots, B_k)'$ is a standard k-dimensional Brownian motion, and $X(0)$ is independent of the Brownian motion.

Under the assumption that μ and $\boldsymbol{\sigma}$ are Lipschitz (or locally Lipschitz), $f(\mathbf{X}(t)) - \int_{[0,t]} Af(\mathbf{X}(s))ds$ is a martingale (local martingale) for every $f \in C_b^2(\mathbb{R}^k)$, the set of all real-valued bounded functions on $\mathbb{R}^k$ having bounded derivatives of first and second order.

In the absence of the smoothness, assumptions made on $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ the stochastic integral in (23.3) may still make sense, but the proof of the existence of a pathwise unique solution may break down. Tanaka's one-dimensional equation

$$dX(t) = \operatorname{sgn} X(t)dB(t),\ X(0) = 0, \tag{23.5}$$

provides a well-known example. Let $W(t)$, $t \ge 0$ be a standard Brownian motion starting at 0. Define the stochastic integral

[1] In addition to that of its originators, Stroock and Varadhan (1979), a more thorough treatment than the present chapter can be found in the text by Ethier and Kurtz (1985).

$$Y(t) = \int_{[0,t]} \mathrm{sgn}(W(s))dW(s), \quad dY(t) = \mathrm{sgn}\, W(t)dW(t). \tag{23.6}$$

By Lévy's Theorem 9.4, because Y has quadratic variation t on $[0, t]$, it is a standard Brownian motion. Note also that $\int_{[0,t]} \mathrm{sgn}(W(s))dY(s) = \int_{[0,t]} \mathrm{sgn}^2(W(s))dW(s) = W(t)$. Thus, W is a solution to Tanaka's SDE:

$$dW(t) = \mathrm{sgn}\, W(t)dY(t), \tag{23.7}$$

but *driven by the Brownian motion Y, not B*. One can of course choose B or W in (23.6) to get B as the solution $dB(t) = \mathrm{sgn}\, B(t)dY(t)$. Apart from the fact that one has to construct a new driving Brownian motion Y, the solution W is *not pathwise unique*. For example, $-W(t)$, $t \geq 0$ is another solution. However, the two solutions have the same law. Although there does not seem to be a universal agreement on nomenclature, we shall call a solution of an SDE which may not be pathwise unique or may not be driven by the underlying Brownian motion, a *weak solution*.

The martingale problem of Stroock and Varadhan (1969,1979) may be stated as follows, generalizing the martingale formulation indicated at the outset (by the semigroup formalism).

Definition 23.1 (Martingale Problem) Let the second order operator A be as in (23.1). Let $C[0, \infty) \equiv C([0, \infty) : \mathbb{R}^k)$ be the set of all continuous $\mathbb{R}^k$-valued functions on $[0, \infty)$, $\mathcal{B}$ the σ-field of all Borel measurable subsets of the metric space $C[0, \infty)$ endowed with the topology of uniform convergence on the bounded intervals of $[0, \infty)$. Denote the canonical filtration $\mathcal{B}_t = \sigma\{\mathbf{x}(u), 0 \leq u, \leq t\}$ where $\mathbf{x}(u)$ is the projection on the uth coordinate of $C[0, \infty)$. Then a solution to the (*local*) *martingale problem for* $(\mathbf{x}, A)$ is a probability P on $(C[0, \infty), \mathcal{B})$ such that $P(\mathbf{x}(0) = \mathbf{x}) = 1$ and for all $f \in C^2(\mathbb{R}^k)$,

$$f(\mathbf{x}(t)) - \int_{[0,t]} Af(\mathbf{x}(s))ds \text{ is a local } \{\mathcal{B}_t\} \text{ martingale under } P. \tag{23.8}$$

Here "local" means $f(\mathbf{x}(t \wedge \tau_n)) - \int_{[0,t\wedge\tau_n]} Af(x(s))ds$ is a $\{\mathcal{B}_t\}$-martingale for all $n \in \mathbb{N}$, $\tau_n := \inf\{t \geq 0 : |\mathbf{x}(t)| \geq n\}$. In more detail one says $(C([0, \infty), \{\mathcal{B}_t\}, \mathcal{B}, P)$ is a solution to the local martingale problem for $(\mathbf{x}, A)$.

Remark 23.2 The lower case notation for the formulation adopted here is merely a mathematical convenience, indicating the use of the canonical model $\Omega = C([0, \infty))$.

Let the coefficients $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ be measurable and locally bounded. It follows from Itô's Lemma that if $\mathbf{X}(t)$, $t \geq 0$ is a weak solution to the SDE (23.2) with $\mathbf{X}(0) = \mathbf{x}$, then its distribution P (on $(C[0, \infty), \mathcal{B})$ is a solution to the local martingale problem $(\mathbf{x}, A)$. The converse is also true. Namely,

Proposition 23.2 *If there exists an augmented solution-space* $(\Omega, \mathcal{F}, P)$*, with a filtration* $\{\mathcal{B}_t\}$ *such that (23.8) holds, then there exists a filtration* $\{\mathcal{F}_t\}$ *and an* $\{\mathcal{F}_t\}$*-adapted Brownian motion* B_t*,* $t \geq 0$ *and a* $\{\mathcal{F}_t\}$*-adapted process* $\mathbf{X}_t$ *such that (23.2) holds.*

Proof To prove this, for simplicity, we assume the diffusion matrix to be nonsingular. First let $\boldsymbol{\sigma}(\cdot)$ be the positive square root of $\mathbf{a}(\cdot)$ (obtained from the spectral decomposition of $\mathbf{a}(\cdot)$). Consider on $(C[0,\infty), \{\mathcal{B}_t\}, \mathcal{B}, P)$,

$$M(t) = \mathbf{x}(t) - \mathbf{x}(0) - \int_{[0,t]} \boldsymbol{\mu}(\mathbf{x}(s))ds, \tag{23.9}$$

where $M(t) = (M^1(t), \ldots, M^k(t))$. It follows from (23.8), with $f(\mathbf{x}) = x_i$ that M^i is a local martingale ($i = 1, \ldots, k$). Also consider the local martingales corresponding to $f(\mathbf{x}) = x_i x_j$ in (23.8). That is,

$$M_t^{(ij)} := x_i(t)x_j(t) - x_i(0)x_j(0) - \int_{[0,t]} [x_j(s)\mu_i(\mathbf{x}(s)) + x_i(s)\mu_j(\mathbf{x}(s)) + a_i j(\mathbf{x}(s))]ds \tag{23.10}$$

is a local martingale for every pair (i, j). This may be expressed as (see Remark 19.5, and Exercise 1)[2]

$$\begin{aligned} M_t^{(ij)} = x_i(0)x_j(0) + x_j(0)x_i(0) &+ \int_{[0,t]} x_j(s)M^i(ds) + \int_{[0,t]} x_i(s)M^j(ds) \\ &+ \langle M^i, M^j\rangle_t - \int_{[0,t]} a_{ij}(\mathbf{x}(s))ds, \end{aligned} \tag{23.11}$$

where $\langle M^i, M^j\rangle_t$ is the quadratic covariation of the martingales M^i, M^j (see Definition 19.7). Then $\langle M^i, M^j\rangle_t - \int_{[0,t]} a_{ij}(\mathbf{x}(s))ds$ is a continuous local martingale of locally finite variation, starting at 0; hence it equals 0 a.s. That is, $\langle M^i, M^j\rangle_t = \int_{[0,t]} a_{ij}(\mathbf{x}(s))ds$, almost surely. Then one has

$$\begin{aligned} W_t &:= \int_{[0,t]} \boldsymbol{\sigma}(\mathbf{x}(s))^{-1} M(ds), \\ W_t^i &= \sum_{1\leq j\leq k} \int_{[0,t]} \sigma^{ij}(\mathbf{x}(s))M^j(ds);\ ((\sigma^{ij}(\cdot))) = \boldsymbol{\sigma}^{-1}(\cdot), \\ \frac{d}{dt}\left\langle W_t^i, W_t^j\right\rangle &= \sum_{1\leq \ell\leq k} \sigma^{i\ell}(x(t))\sigma^{\ell j}(x(t))\frac{d}{dt}\langle M^i, M^j\rangle = \delta_{ij}. \end{aligned}$$

[2] See Scheutzow (2019) for a more general result.

Hence W_t is a standard k-dimensional Brownian motion, $M(dt) = \boldsymbol{\sigma}(x(t))dW_t$, and (23.9) reduces to

$$\mathbf{x}(t) = \mathbf{x}(0) + \int_{[0,t]} \boldsymbol{\mu}(\mathbf{x}(s))ds + \int_{[0,t]} \boldsymbol{\sigma}(\mathbf{x}(s))dW_s.$$

■

Having established the equivalence of the (local) existence of a weak solution to that of a solution to the (local) martingale problem for $(\mathbf{x}, A)$ for every initial $\mathbf{x}$, we briefly address the question of existence of a solution and uniqueness. Itô's theory shows if the coefficients $\boldsymbol{\mu}(\cdot)$ and $\boldsymbol{\sigma}(\cdot)$ are locally Lipschitz, then this solution exists and is (pathwise) unique up to explosion time (see Chapter 12). Note that, in Itô's theory, it is not necessary to assume that $\boldsymbol{\sigma}(\cdot)$ or $\mathbf{a}(\cdot)$ are positive definite. For the relationship between the smoothness of $\boldsymbol{\sigma}(\cdot)$ and that of $\mathbf{a}(\cdot)$, we refer to Friedman (1975, Chapter 6). According to standard PDE theory (see Friedman, loc. cit.), in case (i) $\mathbf{a}(\cdot)$ is uniformly elliptic, i.e., the eigenvalues are bounded away from zero and infinity, (ii) $\boldsymbol{\mu}(\cdot)$ and $\mathbf{a}(\cdot)$ are locally Hölder continuous, there exists a smooth transition probability density of the diffusion with these coefficients satisfying Kolmogorov's backward and forward equations. Of course for the existence and uniqueness of the diffusion (possibly with explosion) with a continuous probability density, it is enough to have the preceding conditions (i) and (ii), hold locally. Using their martingale method and some hard analysis, Stroock and Varadhan (1979) prove the following :

Theorem 23.3 (Stroock and Varadhan) *If $\boldsymbol{\mu}(\cdot)$ is measurable and bounded on compacts, and if the smallest eigenvalue of $\mathbf{a}(\cdot)$ is bounded away from zero on compacts, then there exists a (local) weak solution, which is unique in law for $(\mathbf{x}, A)$ for every $\mathbf{x}$.*

Recall that the uniqueness implies the Markov property of the (weak) solution (see the proof of Theorem 7.2). When the (local) martingale problem for $(\mathbf{x}, A)$ has a solution (for every initial $\mathbf{x}$) that is unique in law, we say that the problem is *well posed*.

The martingale method has been used by Stroock and Varadhan (1979, Theorem 11.2.3) to derive convergence of scaled Markov chains to diffusions[3] Following them, for each fixed $h > 0$, let $p_h(x, dz)$ be the transition probability of a Markov chain $\{X_n\}$ on the lattice $\Delta\mathbb{Z}$, $\Delta = h^{\frac{1}{2}}$, in the canonical model $(\Delta\mathbb{Z}^{\{0,\ldots,\infty\}}, \mathcal{F}_{mh} = \sigma\{x(jh) : j = 0, \ldots, m\})$. By $x(jh)$ we mean the (random) jth iterate of this chain. As in the case of the FCLT (Chapter 17), let P_h^x denote the distribution of the polygonal process on $(C[0, \infty) : \mathbb{R})$ for a Markov chain with transition probability $p_h(x, dy)$ starting at x. Here Ω has the usual topology of uniform convergence on compact intervals of $[0, \infty)$. That is,

[3] See Durrett (1996) Chapter 8, Sec 7, for a nice treatment of convergence for both discrete and continuous parameter Markov chains with interesting examples.

(i) $P_h^x(x(0) = x) = 1$
(ii) $P_h^x[x(t) = \frac{(m+1)h-t}{\Delta} x(mh) + \frac{(t-mh)}{\Delta} x((m+1)h), mh \le t \le (m+1)h] = 1$ for all $m = 0, 1, \dots$
(iii) The conditional distribution of $x((m+1)h)$ given $\mathcal{F}_{mh}$ is $p_h(j\Delta, dy)$ on the set $\{x(\cdot) : x(mh) = j\Delta\}$.

Denote by A_h the linear operator on the space $C_b(\mathbb{R})$ given by

$$A_h f(x) = \int (f(x) - f(z)) p_h(x, dz), \tag{23.12}$$

and note that

$$f(x(jh)) - \sum_{0 \le k \le j-1} A_h f(x(kh)),\ (j = 0, 1, \dots) \text{ is a } \{\mathcal{F}_{jh}\}\text{-martingale} \tag{23.13}$$

under the Markov chain, starting at $x(0)$ or, equivalently, under P_h^x (Exercise 2). Let

$$\begin{aligned} \mu_h &= \frac{1}{h} \int_{\{|z-x| \le 1\}} (z - x) p_h(x, dz) \\ a_h(x) &= \frac{1}{h} \int_{\{|z-x| \le 1\}} (z - x)^2 p_h(x, dz). \end{aligned} \tag{23.14}$$

Proposition 23.4 *Assume that for all $R > 0$ there exist functions $\mu(\cdot)$ and $a(\cdot)$ such that*

(i) $\lim_{h \downarrow 0} \sup_{\{|x| \le R\}} |\mu_h(x) - \mu(x)| = 0,$
(ii) $\lim_{h \downarrow 0} \sup_{\{|x| \le R\}} |a_h(x) - a(x)| = 0$
(iii) $\lim_{h \downarrow 0} \sup_{\{|x| \le R\}} \frac{1}{h} \int |x - z|^3 p_h(x, dz) = 0$

Define the operator $L = \mu(x)\frac{d}{dx} + \frac{1}{2}a(x)\frac{d^2}{dx^2}$. Then for every $f \in C_0^\infty$ (the space of infinitely differentiable functions vanishing at infinity), $\frac{1}{h} A A_h f(x) \to Lf(x)$ as $h \downarrow 0$, uniformly on compact subsets of $\mathbb{R}$.

Proof Consider the Taylor expansion $f(z) = f(x) + f_2(x, z) + r(x, z)$,

$$\begin{aligned} &f_2(x, z) = (z - x) f'(x) + \frac{1}{2}(z - x)^2 f''(x), \\ &|r(x, z)| \le c(f)|z - x|^3, \end{aligned} \tag{23.15}$$

$c(f)$ being a constant depending on f. Let

$$L_h = \mu_h(x) \frac{d}{dx} + \frac{1}{2} a_h(x) \frac{d^2}{dx^2}. \tag{23.16}$$

It is simple to check, using definitions (23.12), (23.14), and (23.16), the Taylor expansion, and (23.4)(iii) that

$$|\frac{1}{h}A_h f(x) - L_h f(x)| \leq \frac{c(f)}{h}\int |z-x|^3 p_h(x,dz) \to 0, \tag{23.17}$$

uniformly on compacts as $h \to 0$. The proposition now follows from the assumptions (23.4)(i)–(iii). ■

Theorem 23.5 *In addition to (23.4)(i)–(iii), assume:*

$$\textit{(iii)':}\ \lim_{h\downarrow 0}\sup_{x\in\mathbb{R}}\frac{1}{h}\int |x-z|^3 p_h(x,dz) = 0.$$

Also, assume

$$\sup_{h\geq 0,\, x\in\mathbb{R}}\{a_h(x) + |\mu_h(x)|\} < \infty \tag{23.18}$$

and that the initial point $x(0) = x_h \to x_0$ as $h \downarrow 0$. Then P^h_{xh} converges weakly to the distribution of the diffusion with drift $\mu(\cdot)$ and diffusion coefficient $a(\cdot)$, starting at x, provided the martingale problem for (x, L) is well posed.

Proof Note that, for any $\varepsilon > 0$,

$$P^h_{xh}(|x(m+h) - x(mh)| \geq \varepsilon) \leq \sup_{x\in\mathbb{R}}\int |x-z|^3 p_h(x,dz) \leq h\cdot o(h),$$

as $h \downarrow 0$. Therefore

$$\sum_{0\leq m\leq T/h} P^{x_h}_h(x(m+h) - x(mh)| \geq \varepsilon) \to 0 \text{ as } h\downarrow 0 \text{ for all } T > 0. \tag{23.19}$$

We now invoke a result adapted to Markov chains (Stroock and Varadhan (1979, Theorem 1.4.11) to conclude that (23.19) implies that $\{P^{xh_n}_{h_n} : n = 1, 2, \ldots\}$ is precompact, for every sequence $h_n \downarrow 0$. In view of Proposition 23.4, it now follows that every limit point is the distribution of a diffusion with drift $\mu(\cdot)$ and diffusion coefficient $a(\cdot)$ starting at x_0 provided the martingale problem for (x, L) is well posed for all x. ■

Example 5 (Birth-and-Death Approximation of Diffusions) Let $\mu(\cdot)$ and $a(\cdot)$ be bounded and Lipschitz, $a(\cdot)$ bounded away from zero. Consider the birth–death chain on $\Delta\mathbb{Z}$, $\Delta = h^{1/2}$, with transition probabilities:

$$p_h(i\Delta, (i-1)\Delta) = \frac{a(i\Delta)}{2a_0} - \frac{\mu(i\Delta)\Delta}{2} =: \delta_i(h)$$
$$p_h(i\Delta, (i+1)\Delta) = \frac{a(i\Delta)}{2a_0} + \frac{\mu(i\Delta)\Delta}{2} =: \beta_i(h),$$
$$p_h(i\Delta, i\Delta) = 1 - \beta_i(h) - \delta_i(h).$$

Here, $a_0 = \sup_{x\in\mathbb{R}} a(x)$. Note that, for sufficiently small $h > 0$, $p_h(i\Delta, j\Delta)$ are probabilities, ($j = i-1, i, i+1$). Then, with $x = i\Delta$,

$$\mu_h(i\Delta) = \frac{1}{h}\int_{\{|z-x|\le 1\}} (z-x)p_h(x, dz) = \mu(i\Delta),$$
$$a_h(i\Delta) = \frac{1}{h}\int_{\{|z-x|\le 1\}} (z-x)^2 p_h(x, dz) = \frac{a(i\Delta)}{a_0}$$
$$\lim_{h\downarrow 0}\sup_{x\in\mathbb{R}} \frac{1}{h}\int |x-z|^3 p_h(x, dz) = 0.$$

For $f \in C_0^\infty$, at $x = i\Delta$,

$$\frac{1}{h}A_h(f(x)) = \frac{1}{h}\int (f(z)-f(x))p_h(x, dz) = \mu(i\Delta)f'(x) + \frac{1}{2}f''(x)\frac{a(i\Delta)}{2a_0} + o(1)$$

as $h \downarrow 0$, uniformly on $\mathbb{R}$.

If we take $a_0 = 1$, then the desired convergence follows. In general, one can use a simple scaling to get the desired result. The boundedness restriction on $a(\cdot)$ and $\mu(\cdot)$ may be relaxed by localization, still requiring that $a(\cdot)$ is bounded away from zero on compacts, and both $\mu(\cdot)$ and $a(\cdot)$ are Lipschitz.

Remark 23.3 As an alternative to the martingale method, one may prove the approximation in Example 5 and also that of the transition probability density of the diffusion, by appealing to numerical analysis, laid out in detail in Bhattacharya and Waymire (1990, 2009), pp. 386–389.

Exercises

1. Verify Equation (23.11).
2. Show that $f(x(jh)) - \sum_{0\le k\le j-1} A_h f(x(kh))$ for $j = 0, 1, \ldots$ is a $\{\mathcal{F}_{jh}\}$-martingale under the Markov chain, starting at $x(0)$ (see Equation (23.13)).
3. Suppose that P is a unique solution to the martingale problem for $(\mathbf{x}, A)$. Let $\psi_t(B) = P(\{x \in C_0[0\infty) : \mathbf{x}(t) \in B\}), B \in \mathcal{B}(\mathbb{R}^k), t \ge 0$. (i) Show that $\{\psi_t : t \ge 0\}$ is a (weak) solution to the forward equation in the sense that $\int_{\mathbb{R}^k} f(\mathbf{y})\psi_t(d\mathbf{y}) = \int_{\mathbb{R}^k} f(\mathbf{y})\psi_0(d\mathbf{x}) + \int_{[0,t]} \psi_s(\mathbf{y})Af(\mathbf{y})ds,\ f \in C_0^2(\mathbb{R}^k), t \ge 0.$

[*Hint*: Take expected values in $M(t)$.] (ii) Suppose that $\psi_t(d\mathbf{y}) = \varphi(t, \mathbf{y})d\mathbf{y}, t > 0$ has a twice continuously differentiable density $\varphi(t, \mathbf{y})$ with respect to Lebesgue measure. Verify Kolmogorov's forward equation for $f \in C_0^2(\mathbb{R}^k)$:

$$\frac{\partial \varphi}{\partial t}(t, \mathbf{y}) = \int_{\mathbb{R}^k} [\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{\partial^2 (a_{ij}(\mathbf{y})\varphi(t, \mathbf{y}))}{\partial y^{(i)} \partial y^{(j)}} - \sum_{i=1}^{k} \frac{\partial(\mu^{(i)}(\mathbf{y})\varphi(t, \mathbf{y}))}{\partial y^{(i)}}] f(y) dy.$$

[*Hint*: Use integration-by-parts as in Chapter 15.]

Chapter 24
Special Topic: Multiphase Homogenization for Transport in Periodic Media

There have been numerous studies, especially in the hydrology literature, mostly empirical, showing that the diffusivity of a solute concentration immersed in a fluid flow increases with spatial scale, sometimes at a rather large rate. This serves to motivate the mathematical theory presented in this chapter. Striking examples are provided to demonstrate the effect of scale as one successively passes from small to larger scales. The main results in this chapter are based on Bhattacharya and Goetze (Bernoulli 1:81–123, 1995) and Bhattacharya (Ann Appl Probab 951–1020, 1999).

Thermal and electric conduction in composite media and diffusion of matter through them are problems of great importance.[1] Examples of such composite media are natural heterogeneous material such as soils, polycrystals, wood, animal and plant tissue, fiber composites, etc. One of the applications of interest in this chapter is the transport of solutes, such as pollutants, in an aquifer. The concentration $c(t, y)$ of matter at time t at a point y in such problems satisfies a Fokker-Planck equation:

$$\partial c(t,y)/\partial t = \frac{1}{2}\sum_{1\le i,j\le k}\frac{\partial^2}{\partial y_i\partial y_j}(D_{ij}(y)c(t,y)) - \sum_{1\le i\le k}\frac{\partial}{\partial y_i}(v_i(y)c(t,y)),$$
$$c(0,dy) = \delta_x(dy). \tag{24.1}$$

[1] See Ben Arous and Owhadi (2003), Bensoussan et al. (1978), and Fried and Combarnous (1971).

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_24

Here $D(\cdot) = ((D_{ij}(\cdot)))_{1\le i,j\le k}$ is a $k \times k$ positive definite matrix-valued function depending on the local properties of the medium, and $v(\cdot)$ is a vector field which arises from other sources. Throughout the spatial dimension is $k > 1$. The initial concentration may be more general than a point source at some point x. One may also think of (24.1) as the equation of transport, or diffusion, of a substance in a turbulent fluid. Indeed, in a classic study, Richardson (1926) analyzed data on atmospheric diffusion over 12 or more different orders of spatial scale and conjectured the relationship

$$D_L \propto L(t)^{4/3}, \quad D_L(t) := L(t)^2/t, \tag{24.2}$$

where D_L is the diffusivity at spatial scale L and $L(t)$ is the root-mean squared distance from the mean flow. To be specific, *diffusivity* is dispersion per unit time, or the expected squared distance traveled in a unit of time due to dispersion. This relationship was later shown by Batchelor (1952) and Obukhov (1941) to be consistent with Kolmogorov's dimensional analysis showing that the velocity spectrum of turbulent fluid is proportional to $L^{1/3}$. The velocity v we consider is smooth and not turbulent. There have been numerous studies, especially in the hydrology literature, mostly empirical, (see, e.g., Fried and Combarnous (1971)), showing that the diffusivity of a solute concentration increases with spatial scale, sometimes at a rather large rate. Some theoretical studies such as Winter et al. (1984), Gelhar and Axness (1983), have looked at the dispersion assuming that the velocity v is an ergodic random field and analyzed the relationship between the correlation structure and distance. The present study provides, in particular, the relationship between $L(t)$ and $D_L(t)$ at different spatial scales for the transport of a substance, under a smooth velocity v composed of a local component b and a large-scale component β:

$$v(y) = b(y) + \beta(y/a). \tag{24.3}$$

Leaving aside the beginning part (Theorems 24.1, 24.2), b and β, and therefore v are assumed to be divergence-free, i.e., $\mathrm{div}v = 0$, thereby preventing local accumulation, and for the most part, b and β are assumed to be (spatially) periodic. The spatial parameter "a" is assumed to be "large" compared to the local scale, and precise asymptotics in law are provided over times t in relation to "a."

Mathematically, or even physically, (24.1) represents the diffusion of solute particles in a composite medium governed by the stochastic differential equation:

$$dX(t) = [b(X(t)) + \beta(X(t)/a)]dt + \sigma(X(t))dB(t), \quad X(0) = 0, \tag{24.4}$$

where $\sigma(\cdot)$ is the positive square root of $D(\cdot)$ and $\{B(t) : t \ge 0\}$ is a standard k-dimensional Brownian motion. The velocity v is assumed to be of the form (24.3). We write $x = (x_1, \ldots, x_k)$, $b = (b_1, \ldots, b_k)$, $\beta = (\beta_1, \ldots \beta_k)$.

We begin with two interesting examples to illustrate the theory underlying the evolution of concentration through Gaussian and non-Gaussian phases and different diffusivities as time progresses.

Example 1 Let $k = 2$ and $\{B(t) = (B_1(t), B_2(t)) : t \geq 0\}$ a two-dimensional standard Brownian motion. Consider the diffusion process $\{X(t) = (X_1(t), X_2(t)) : t \geq 0\}$, defined by

$$\begin{aligned} dX_1(t) &= [c_1 + c_2 \sin(2\pi X_2(t)) + c_3 \cos(2\pi X_2(t))/a)]dt + dB_1(t), \\ dX_2(t) &= dB_2(t), \quad X(0) = (0, 0). \end{aligned} \tag{24.5}$$

The constants $c_j (j = 1, 2, 3)$ are all nonzero.

Here N denotes the Normal distribution, $\mathcal{L}$ stands for law or distribution, $\Rightarrow$ denotes convergence in law, and $\Rightarrow \mathcal{L}(Z)$ denotes convergence in law to the distribution of a random variable Z. Also, the ratio of two sides of $\sim$ converges to one. In this example,

$$k = 2, b_1(y) = c_1 + c_2 \sin(2\pi y_2), \quad \beta_1(y) = c_3 \cos(2\pi y_2/a),$$

while $b_2(y) = \beta_2(y) = 0$. One could make it a k-dimensional example, $k > 2$, by letting $b_j(y) = \beta_j(y) = 0$, for $2 < j \leq k$, with noise $dB_j(t)$. Then $X_j(t), 2 < j \leq k$, would just be independent Brownian motions, with diffusivity one in these coordinates.

This is an example of *stratified media*, with the growth in diffusivity only in the first coordinate. Table 24.1 displays this growth with time. The first line represents an asymptotic Normal distribution around the mean flow. In this range of time the large-scale effect can be ignored. Due to periodicity,

$$Z(t) := X(t) \bmod 2\pi$$

is an ergodic diffusion on $\mathbb{T}_1$, the unit two-dimensional torus, with the uniform invariant distribution. Hence the mean flow, namely, mean of $b(Z(t))$, over time t converges to the phase average $\widetilde{b}$ of $b(\cdot)$ with respect to the uniform distribution on the unit torus. The stochastic integral is then the integral of a martingale with stationary ergodic increments and, when divided by $\sqrt{t}$, converges to the Normal

Table 24.1 Scale regimes for Example 1.

Time scale	Asymptotic law
$1 << t << a^{4/3}$	$[X_1(t) - (c_1 + c_3)t]/\sqrt{t} \Rightarrow N(0, 1 + c_2^2/2\pi^2)$
$a^{4/3} << t << a^2$	$[X_1(t) - (c_1 + c_3)t]/(t^2/a^2) \Rightarrow \mathcal{L}(2c_3\pi^2 \int_{[0,1]} B_2^2(s)ds)$
$t/a^2 \sim r \neq 0$	$[X_1(t) - c_3 t]/t \Rightarrow \mathcal{L}((c_3/r) \int_{[0,r]} \cos(2\pi B_2(s))ds)$
$t >> a^2$	$[X_1(t) - c_3 t]/a\sqrt{t} \Rightarrow N(0, c_3^2/2\pi^2)$

distribution with dispersion computed by the formula derived in Chapter 17. The diffusivity is larger than what it would be without the drift $b(\cdot)$ but only by a constant. As time becomes larger than the initial range, namely, $O(a^{4/3})$, the asymptotics dramatically change. In lines 2 and 3 in Table 24.1, it is shown that the Normal distribution breaks down, and the local velocity only contributes to the mean flow, while the large-scale fluctuations $\beta(\cdot/a)$ causes an increase in the order of magnitude of diffusivity. In line 3, $D_L \sim L$, the maximum diffusivity. In line 4, when $t >> a^2$, the diffusion is again asymptotically normal but with a large diffusivity. After this the diffusion settles down to this final Gaussian phase with diffusivity gradually going down, in the absence of any further source of heterogeneity. If one expresses the relationship between L and D_L as

$$D_L \propto L^{\theta}, \tag{24.6}$$

then θ grows from 0 to 1 with time and then starts going down to 0 after the final phase (Exercise 11). Note that if in the final phase one lets $t \sim a^{2+\gamma}$ for some $\gamma \geq 0$, then

$$D_L \sim a^2 \sim t^{\frac{2}{2+\gamma}}.$$

Therefore, θ attains the maximum value 1 at $\gamma = 0$ and decreases as γ increases. The range in the first line, namely, $1 << t << a^{4/3}$, for stratified media, is checked in Theorem 24.1. The ranges and computations in lines 2 and 3 may be carried out directly, using periodicity, Itô's Lemma, martingale CLT, etc. That is, they do not require a special general theory (see Bhattacharya (1999), pp. 3003–3005). But the derivation of the final phase asymptotics, in line 4, requires a special theory that will be presented in this chapter (see Theorem 24.3).

The significance of this theory is further manifested in the following Example 2, which at first appearance does not seem to differ much from Example 1; but it has profoundly different asymptotics. In particular, in this example the diffusivity exponent does not grow from 0 i.e., $\theta = 0$ for any range of t! (see Theorem 24.2).

Example 2 As in Example 1, consider $k = 2$. $X_1(t)$ in this example is exactly the same as in Example 1. But one applies a constant drift in $X_2(t)$:

$$\begin{aligned} dX_1(t) &= [c_1 + c_2 \sin(2\pi X_2(t)) + c_3 \cos(2\pi X_2(t))/a)]dt + dB_1(t), \\ dX_2(t) &= \delta dt + dB_2(t), \quad \delta \neq 0,\, X(0) = 0 = (0,0). \end{aligned} \tag{24.7}$$

For the first two rows of Table 24.2, the Gaussian convergence follows arguments based on asymptotics of diffusions which are integrals of ergodic Markov processes

$$Z(t) := X(t) \bmod 2\pi$$

Table 24.2 Scale regimes for Example 2.

Time scale	Asymptotic law
$1 << t << a^{2/3}$	$[X_1(t) - (c_1 + c_3)t]/\sqrt{t} \Rightarrow N(0, 1 + c_2^2/[2(\delta^2 + \pi^2)]$
$t/a^{2/3} \to r > 0$	$[X_1(t) - (c_1 + c_3)t]/\sqrt{t} \Rightarrow N(0, 1 + c_2^2/[2(\delta^2 + \pi^2)]$
$t >> a^2$	$[X_1(t) - c_1 t]/\sqrt{t} \Rightarrow N(0, 1 + c_2^2/[2(\delta^2 + \pi^2)] + c_3^2/2\delta^2)$

as in Example 1. There is a big gap in time scales between the second row and the third row. But it seems very likely that the diffusivity during this period is no more than that in the final phase. What would appear very surprising is that in the final phase $t >> a^2$, the diffusivity does not really grow, except for a constant term. Theorem 24.3 analyzes the precise mathematical distinction between the two cases.

First Phase Asymptotics Without essential loss of generality, we may take as the initial point $x = 0$. This essentially means that the concentration spreads starting with a point mass injection at a point we label 0. One may assume a somewhat more general concentration profile which has not spread very far. Let $\{X(t) : t \geq 0\}$ be as defined in (24.5). We use constants c_i, c_i', d_i, etc., independent of t and a.

Theorem 24.1 *Let $x_0 = 0$. Assume that $b(\cdot)$ and its first order derivatives are bounded, and $\beta(\cdot)$, $D(\cdot)$ and their first and second-order derivatives are bounded. Let $X(t)$ be as in (24.4), and let $Y(t)$ be defined by*

$$dY(t) = [b(Y(t)) + \beta(0)]dt + \sigma(Y(t))dB(t), \quad Y(0) = 0, \tag{24.8}$$

which ignores the large-scale variability. Let P_t be the distribution of $\{X(s) : 0 \leq s \leq t\}$ and Q_t the distribution of $\{Y(s) : 0 \leq s \leq t\}$. (i) Then there exist constants $d_j (j = 1, \dots, 5)$ such that for $a, t > 1$,

$$\begin{aligned} &||P_t - Q_t||_{TV} \\ &\leq d_1 \frac{t^{3/2}}{a^2} + d_2 \frac{t^{3/2}}{a} + d_3 \frac{t}{a} \leq d_4 \frac{t^{3/2}}{a^2} + d_2 \frac{t^{3/2}}{a} \leq d_5 \frac{t^{3/2}}{a}. \end{aligned} \tag{24.9}$$

Here $||\cdot||_{TV}$ is total variation norm. (ii) In the special case $b_j = 0$ for $2 \leq j \leq k$, $\beta_j(0) = 0$ for $2 \leq j \leq k$ and $\frac{\partial \beta_j}{\partial x_1} = 0$ for $1 \leq j \leq k$, one has $d_2 = 0$, so that

$$||P_t - Q_t||_{TV} \leq d_1 \frac{t^{3/2}}{a^2} + d_3 \frac{t}{a}. \tag{24.10}$$

If, in addition, as in Example 1, $\nabla \beta(0) = 0$, then one may take $d_3 = 0$ in (24.10).

Proof By the Cameron–Martin–Girsanov Theorem (see Theorem 9.9), the Radon-Nikodym derivative of P_t with respect to Q_t is given by e^{V_t}, where

$$V_t = \int_{[0,t]} \sigma^{-1}(Y(s))\{\beta(Y(s)/a) - \beta(Y(0)/a)\}dB(s)$$
$$-(1/2)\int_{[0,t]} |\sigma^{-1}(Y(s))\{\beta(Y(s)/a) - \beta(Y(0)/a)\}|^2 ds.$$

Since $\{e^{V_t} : t \geq 0\}$ is a martingale, $\mathbb{E}e^{V_t} = 1$, and

$$0 = \mathbb{E}\left(1 - e^{V_t}\right) = \mathbb{E}\left(1 - e^{V_t}\right)^+ - \mathbb{E}\left(1 - e^{V_t}\right)^-.$$

Also $(1 - e^x)^+ = 0$ if $x > 0$, and $1 - e^x \leq |x| \wedge 1$ for $x \leq 0$. Therefore,

$$||P_t - Q_t||_{TV} = \mathbb{E}|1 - e^{V_t}| = 2\mathbb{E}(1 - e^{V_t})^+ \leq 2\mathbb{E}(|V_t| \wedge 1). \tag{24.11}$$

Now

$$\mathbb{E}|V_t| \leq I_1(t) + I_2(t),$$

where denoting the set of eigenvalues of

$$D(y) := \sigma(y)\sigma'(y),$$

by $\rho(D(y))$, and

$$\lambda = \inf_{y \in \mathbb{R}^k} \rho(D(y)),$$

one has (see (7.42)),

$$I_1(t)^2 = \mathbb{E}|\int_{[0,t]} \sigma^{-1}(Y(s))\{\beta(Y(s)/a) - \beta(Y(0)/a)\}dB(s)|^2$$
$$= \mathbb{E}\int_{[0,t]} |\sigma^{-1}(Y(s))\{\beta(Y(s)/a) - \beta(Y(0)/a)\}ds|^2$$
$$\leq (1/\lambda)\int_{[0,t]} \mathbb{E}|\{\beta(Y(s)/a) - \beta(Y(0)/a)\}ds|^2,$$
$$I_2(t) = (1/2)\int_{[0,t]} \mathbb{E}|\sigma^{-1}(Y(s))\{\beta(Y(s)/a) - \beta(Y(0)/a)\}|^2 ds$$
$$\leq (1/2\lambda)\int_{[0,t]} \mathbb{E}|\{\beta(Y(s)/a) - \beta(Y(0)/a)\}ds|^2. \tag{24.12}$$

By Itô's Lemma (Theorem 8.4), denoting

$$\nabla = \text{grad},$$

one has

$$\beta_j(Y(s)/a) - \beta_j(Y(0)/a)$$
$$= \int_{[0,s]} L_0\beta_j(\cdot/a)Y(s')ds' + \int_{[0,s]} \nabla\beta_j(\cdot/a)(Y(s'))\cdot\sigma(Y(s'))dB(s'), \quad (24.13)$$

where L_0 is the generator of the diffusion $\{Y(t) : t \geq 0\}$, and recalling $D(y) = ((D_{ii'}(y))) = \sigma(y)\sigma'(y)$,

$$L_0 = (b(\cdot) + \beta(0))\cdot\nabla + \frac{1}{2}\sum_{1\leq i,,i'\leq k} D_{ii'}(\cdot)\frac{\partial^2}{\partial y_i\partial y_i'},$$

$$L_0(\beta_j(\cdot)/a)(Y(s')$$
$$= (1/2a^2)\sum_{1\leq i,i'\leq k} D_{ii'}(Y(s')\frac{\partial^2\beta_j}{\partial y_i\partial y_i'}(Y(s')/a$$
$$+(1/a)(b(Y(s')) + \beta(0))\cdot\nabla\beta_j(\cdot)(Y(s')/a). \quad (24.14)$$

Therefore, denoting the Riemann integral on the right side of (24.13) by $J_{1j}(s)$ and the stochastic integral by $J_{2j}(s)$, one has

$$\mathbb{E}(\beta_j(Y(s)/a) - \beta_j(Y(0)/a))^2 \leq 2\mathbb{E}J_{1j}(s)^2 + 2\mathbb{E}J_{2j}(s)^2. \quad (24.15)$$

Letting $||\cdot||_\infty$ denote the supremum of the Euclidean norm of a real-, vector- or matrix-valued function on $\mathbb{R}^k$, we have

$$\mathbb{E}J_{1j}(s)^2$$
$$\leq s^2[(c_1'/a^4)||D||_\infty^2||\max_{i,i'}\frac{\partial^2\beta_j}{\partial y_i\partial y_i'}(\cdot)||_\infty^2 + (2/a^2)s^2||(b(\cdot)+\beta(0))\cdot\nabla\beta_j(\cdot)||_\infty^2],$$

$$\mathbb{E}J_{2j}(s)^2 \leq (s/a^2)||D(\cdot)||_\infty||\nabla\beta_j(\cdot)||_\infty^2. \quad (24.16)$$

Thus

$$\int_{[0,t]}\sum_{1\leq j\leq k}(\mathbb{E}J_{1j}(s)^2 + \mathbb{E}J_{2j}(s)^2)ds$$
$$\leq (t^3/a^4)c_2'||D||_\infty^2\sum_{1\leq j\leq k}||\max_{i,i'}\frac{\partial^2\beta_j}{\partial y_i\partial y_i'}(\cdot)||_\infty^2$$

$$+(2t^3/3a^2) \sum_{1 \le j \le k} ||b(\cdot) + \beta(0)) \cdot \nabla\beta_j(\cdot)||_\infty^2$$

$$+(t^2/2a^2)||D(\cdot)||_\infty \sum_{1 \le j \le k} ||\nabla\beta_j(\cdot)||_\infty^2 = I(t), \tag{24.17}$$

say. Hence, noting (24.12), (24.15)–(24.17), we get

$$\mathbb{E}|V_t| \le I_1(t) + I_2(t) \le (1/\sqrt{\lambda})\sqrt{I(t)} + (1/2\lambda)I(t). \tag{24.18}$$

From this it follows (see Exercise 1) that

$$\mathbb{E}(|V_t| \wedge 1) \le c_3'\sqrt{I(t)}.$$

Thus (Exercise 1(ii))

$$\mathbb{E}(|V_t| \wedge 1) \le c_3'\sqrt{I(t)} \le d_1 t^{3/2}/a^2 + d_2 t^{3/2}/a + d_3(t/a). \tag{24.19}$$

For part (ii), notice that, under the added conditions, the middle term in the right side of (24.17) and, therefore, in (24.19) vanishes. ■

Theorem 24.1 leads to a first phase asymptotic Gaussian law, where $X(\cdot)$ can be approximated by $Y(\cdot)$. We will assume that the local velocity $b(\cdot)$ and the non-singular matrix $\sigma(\cdot)$ are differentiable and periodic with period one in each coordinate, i.e., with period lattice $\mathbb{Z}^k$. Then the process

$$Z(t) := Y(t) \bmod 1$$

is a diffusion on the unit torus

$$\mathbb{T}_1 = \{(y_1 \bmod 1, y_2 \bmod 1, \dots, y_k \bmod 1) : y \in \mathbb{R}^k\}.$$

We denote its generator also by L_0 (see (24.14)) but restricted to periodic functions with period one. It has been shown in Chapter 17 that the functional central limit theorem holds for the process $Y(\cdot)$ (Theorem 17.8), centered at $(\widetilde{b}+\beta(0))t$, and it is asymptotically Gaussian having limiting dispersion matrix K, where $\widetilde{b}$ is the mean of $b(Z(t))$ with respect to the invariant distribution $\pi(y)dy$ of $Z(t)$,

$$\widetilde{b} = \int \mathbb{T}_1 b(y)\pi(y)dy = (\widetilde{b}_1, \dots, \widetilde{b}_k),$$

$$K = \int_{\mathbb{T}_1} (\mathrm{grad}\psi(x) - I_k)D(x)(\mathrm{grad}\psi(x) - I_k)^T \pi(x)dx, \tag{24.20}$$

where the components of $\psi(y) = (\psi_1(y), \dots, \psi_k(y))$ are unique mean zero (periodic) solutions of

$$L_0\psi_j(y) = b_j(y) - \widetilde{b}_j, \quad j = 1, \ldots, k. \tag{24.21}$$

Also I_k is the $k \times k$ identity matrix. We thus have the following result.

Theorem 24.2 *Assume that $b(\cdot)$ is continuously differentiable, $\sigma(\cdot)$ is Lipschitzian, and both are periodic of period one in each coordinate, as is the continuously differentiable matrix σ. Also assume that β has continuous and bounded first and second derivatives.*

a. *Then in the limit as*

$$n \to \infty, a \to \infty, \text{ such that } n/a^{2/3} \to 0, \tag{24.22}$$

the process

$$[X(nt) - nt(\widetilde{b} + \beta(0))]/\sqrt{n}, \quad t \geq 0, \tag{24.23}$$

starting at 0, converges in law to a k-dimensional Brownian motion with a dispersion matrix K given by (24.20).

b. *Under the additional hypothesis in Theorem 24.1(ii), the convergence holds over the wider time scale $1 << t << a^{4/3}$, i.e., by replacing (24.20) by*

$$n/a^{4/3} \to 0 \tag{24.24}$$

For a proof we refer to Theorem 17.8 in Chapter 17. The central limit theorem for diffusions with periodic coefficients was proved by Bensoussan et al. (1978). Also see, for the present version, Bhattacharya (1985).

Example 3 (Example 1 Asymptotics) In Examples 1 and 2, the parameters of the Gaussian law of Theorem 24.2 can be computed directly without referring to more general results in Chapter 17, or using the expression for K in (24.20) (Exercise 2). This takes care of the convergence to the Gaussian law in the first row of both examples, since the X-process can be approximated by the Y-process in variation norm on $[0, t]$. What happens between the first phase and the last phase time zones depends on specifics of the coefficients. Derivations of the asymptotics in the intermediate phases for Example 1 are indicated in Exercises 3, 4.

For the final phase in Table 24.1, row 4, since $a\sqrt{t} >> \sqrt{t}$, one only needs to evaluate the asymptotic distribution of

$$c_3/(a\sqrt{t}) \int_{[0,t]} \cos(2\pi B_2(s)/a)ds. \tag{24.25}$$

For this phase, we take "a" to be a positive integer. Since the function

$$f(y) = -(c_3a^2/2\pi^2)\cos(2\pi y/a)$$

satisfies

$$\frac{1}{2}f''(y) = c_3 \cos(2\pi y/a),$$

Itô's lemma shows that (24.25) equals

$$\begin{aligned}
&(1/a\sqrt{t})(c_3 a^2/2\pi^2)[\cos(2\pi B_2(t)/a) - 1] \\
&\quad + (1/a\sqrt{t})) \int_{[0,t]} (c_3 a/\pi)\sin(2\pi B_2(s)/a)dB_2(s) \\
&\approx (c_3/\pi\sqrt{t}) \int_{[0,t]} \sin(2\pi B_2(s)/a)dB_2(s),
\end{aligned} \tag{24.26}$$

where $\approx$ indicates that the two sides differ by a quantity which converges to zero. Note that the first term on the left in (24.26) goes to zero, since $t >> a^2$. Now the stochastic integraI $I(t)$ is a martingale whose quadratic variation is

$$\begin{aligned}
Q(t) &= (c_3^2/\pi^2)(1/t) \int_{[0,t]} \sin^2(2\pi B_2(s)/a)ds \\
&=^{\mathcal{L}} (c_3^2/\pi^2)(1/t) \int_{[0,t]} \sin^2(2\pi B_2(s/a^2)ds \\
&= (c_3^2/\pi^2)(1/t/a^2) \int_{[0,t/a^2]} \sin^2(2\pi B_2(u))du \\
&\to (c_3^2/\pi^2) \int_{[0,1]} \sin^2(2\pi y)dy = (c_3^2/2\pi^2),
\end{aligned} \tag{24.27}$$

almost surely, as $t/a^2 \to \infty$, by the ergodic theorem (Exercise 6). Note that $2\pi B_2(u)$ mod1 is a diffusion on the unit torus which converges in variation norm to the invariant probability, namely, the uniform distribution, exponentially fast as $u \to \infty$, uniformly with respect to the initial distribution. Thus (24.27) implies the desired result, via the martingale CLT (see, e.g., Bhattacharya and Waymire (2022), Theorem 15.5) (Exercise 6). Another way to derive the convergence to the Gaussian law is via time change using stochastic integrals (see Corollary 9.5) (Exercise 8).

In Example 2, the computations for the middle and last rows of Table 24.2 are a little more elaborate, as we show now.

Example 4 (Example 2 Asymptotics) As pointed out above, the convergence in law to the Gaussian in the first row of Table 24.2 follows by direct computation. Here we present the computations for the intermediate phase in row 2 in some detail. As before, we let $X(0) = 0$. One then has

$$\frac{X_1(t) - t(c_1 + c_3)}{\sqrt{t}}$$
$$= c_2 \frac{1}{\sqrt{t}} \int_{[0,t]} \sin(2\pi Z(s))ds + c_3 \frac{1}{\sqrt{t}} \int_{[0,t]} [\cos(2\pi Z(s))/a)] - 1]ds + \frac{1}{\sqrt{t}} B_1(t), \tag{24.28}$$

where $X_2(s) = B_2(s) + s\delta$ and $Z(s) = X_2(s) \bmod 1$ is a diffusion on the unit two-dimensional torus $\mathbb{T}_1$. Now $\{Z(s) : s \geq 0\}$ is a Markov process on the one-dimensional torus $[0, 1]$, with 0 and 1 identified and the distribution of $Z(s)$ converges exponentially fast in variation norm, as $s \to \infty$, to its invariant probability, namely, the uniform distribution. Note also that, under the invariant distribution,

$$\mathbb{E} \sin(2\pi Z(s)) = \int_{[0,1]} \sin(2\pi y)dy = 0.$$

It follows from Theorem 17.1 that

$$(1/\sqrt{t}) \int_{[0,t]} \sin(2\pi Z(s))ds \Rightarrow N(0, \nu), \ \ \nu = -2 \int_{[0,1]} (\sin 2\pi y)u(y)dy,$$

where $u(y)$ is the mean-zero solution of

$$\frac{1}{2} u''(y) + \delta u'(y) = \sin 2\pi y.$$

A direct computation shows

$$u(y) = -(2\delta/(\pi^2 + \delta^2)[(\cos 2\pi y)/2\pi + \sin 2\pi y/2\delta].$$

Plugging this in the expression for ν, one gets

$$\nu = (2\delta/(\pi^2 + \delta^2))(1/2\delta) \int_{[0,1]} (\sin^2 2\pi y)dy = 1/(2(\pi^2 + \delta^2).$$

With the contribution from the independent last term in (24.28), one arrives at the desired result, if the middle term on the right in (24.25) goes to zero. Proof of this is indicated in Exercise 5. We now turn to the asymptotics for $t >> a^2$. Use Itô's Lemma to write

$$\frac{X_1(t) - X_1(0) - c_1 t}{\sqrt{t}} = \frac{w(X_2(t)) - w(X_2(0))}{\sqrt{t}} - \frac{1}{\sqrt{t}} \int_{[0,t]} w'(X_2(s))dB_2(s)) + \frac{B_1(t)}{\sqrt{t}}, \tag{24.29}$$

where w is the mean-zero (under the invariant probability of $Z(t)$) solution of

$$\frac{1}{2} w''(y) + \delta w'(y) = c_2 \sin 2\pi y + c_3 \cos(2\pi y/a).$$

By a direct computation, one obtains

$$w(y) = -(c_2\delta/(\pi^2 + \delta^2))[(\cos 2\pi y)/2\pi + (\sin 2\pi y)/2\delta]$$
$$+(c_2\delta a^3/(\delta^2 a^2 + \pi^2)[\sin(2\pi y/a)/2\pi - \cos(2\pi y/a)/2\delta a],$$

$$\begin{aligned} w'(y) &= -(c_2\delta/(\pi^2 + \delta^2))[(\sin 2\pi y) + \pi(\cos 2\pi y)/\delta] \\ &+(c_2\delta a^2/(\pi^2 + \delta^2 a^2)[\cos(2\pi y/a) + (\pi/\delta a)\sin(2\pi y/a)] \\ &= I_1(y) + I_2(y), \end{aligned}$$

say. Since w is bounded, the first term on the right in (24.29) can be neglected. Also, omitting the last term in $w'(y)$ above, which is of the order $1/a$, the term involving the stochastic integral on the right side of (24.29) is a martingale

$$(1/\sqrt{t})\int_{[0,t]}(I_1(X_2(s) + I_2(X_2(s))dB_2(s),$$

and its quadratic variation is

$$\frac{1}{t}\int_{[0,t]}\left[I_1^2(X_2(s)) + I_2^2(X_2(s)) + 2I_1(X_2(s))I_2(X_2(s))\right]ds.$$

As argued in the case (24.27),

$$\frac{1}{t}\int_{[0,t]} I_1^2(X_2(s))ds \to c_2^2/2(\pi^2 + \delta^2) \text{ as } t \to \infty.$$

Also,

$$\begin{aligned} \frac{1}{t}\int_{[0,t]} I_2^2(X_2(s)) &= (c_3/\delta)^2\frac{1}{t}\int_{[0,t]}\cos^2(2\pi(B_2(s)/a + s\delta/a))ds \\ &= \mathcal{L}(c_3/\delta)^2(1/t))\int_{[0,t]}\cos^2(2\pi(B_2(s/a^2) + s\delta/a))ds \\ &= (c_3/\delta)^2(1/(t/a^2))\int_{[0,t/a^2]}\cos^2(2\pi(B_2(s') + as'\delta))ds' \\ &\to c_3^2/2\delta^2 \text{ in probability as } t/a^2 \to \infty. \end{aligned}$$

For the Brownian motion $\{2\pi B_2(u) \bmod 1 : u \geq 0\}$, on the unit circle converges to the uniform distribution in total variation, uniformly over all initial states, and $\int_{[0,1]}(\cos^2 2\pi y)dy = 1/2$. Thus one may use the ergodic theorem to get the limit. We will next show that the integral of the product term in the quadratic variation goes to zero as $t/a^2 \to \infty$. For this express it as

$$\frac{2}{t}\int_{[0,t]}[\sin 2\pi(B_2(s)+s\delta)\cos 2\pi(B_2(s)/a+s\delta/a)]ds$$
$$=^{\mathcal{L}} 2/(t/a^2)\int_{[0,t/a^2]}\sin 2\pi(aB_2(s)+a^2s\delta)\cos 2\pi(B_2(s)+as\delta)ds$$
$$=2/(t/a^2)\int_{[0,t/a^2]}\sin 2\pi(aZ_2(s))\cos 2\pi(Z_2(s))ds,$$

where $Z_2(s)=(B_2(s)+as\delta)\bmod 1$. Since, as argued earlier, $Z(s):=(B_2(s)+y)\bmod 1$ is an ergodic diffusion on the unit circle approaching its invariant probability (uniform distribution) exponentially fast in total variation distance, uniformly in y, as $s\to\infty$. The last expression in the above display equals

$$(2/ta^2)\int_{[0,t/a^2]}[\sin 2\pi((a+1)Z(s))+\sin 2\pi((a-1)Z(s))]ds.$$

Now, uniformly in "a" (positive integer),

$$(1/A)\int_{[0,A]}[\sin(2\pi(a+1)Z(s)))ds\to 0 \text{ in probability} \tag{24.30}$$

as $A\to\infty$ (Exercise 9). Finally, the last term on the right in (24.29), namely, $B_1(t)/\sqrt{t}$ is independent of the other terms and has the standard normal distribution with mean zero and variance one. Collecting the terms together, one gets a Gaussian distribution with mean zero and having variance $1+c_2^2/2(\pi^2+\delta^2)+c_3^2/2\delta^2$.

Final Phase Asymptotics For the analysis of the limiting properties of $\{X(t):t\geq 0\}$, the following assumptions will be made:

- A1 $b(\cdot)$ and $\beta(\cdot)$ are continuously differentiable and periodic with period one (in each coordinate), i.e., with period lattice $\mathbb{Z}^k$. The dispersion matrix D of $X(\cdot)$ is a constant positive definite matrix, $D=\sigma\sigma'$.
- A2 $\mathrm{div}b(\cdot)=0=\mathrm{div}\beta(\cdot)$.
- A3 "a" is a positive integer.

Remark 24.1 In A3 it is enough to require that "a" is rational: $a=p/q$, with positive integers p and q having no common factor other than 1. One may then take the unit of spatial scale as $(1/q)$-th of the original unit, in which case p takes the place of "a." From a physical point of view, one may consider "a" to be a large (fixed) number and the unit of time to be that needed by the mean flow to travel one unit of distance.

Remark 24.2 One may take the matrix σ, and therefore D, to be spatially periodic with period lattice $\mathbb{Z}^k$. Asymptotic results are still valid, although computations become more complicated. However, because of the first phase result, namely, Theorem 24.2, for the final phase, one may assume σ and D to be constant matrices, without any essential loss of generality for the problem at hand.

Remark 24.3 It is not necessary that the coefficients, $a(\cdot)$, $\beta(\cdot)$ have period one, i.e., a period lattice $\mathbb{Z}^k$. Any common period lattice suffices, because a linear map can transform the lattice to $\mathbb{Z}^k$.

We begin with the fact mentioned earlier (see (24.20) and (24.21) and the paragraph preceding them) that, for a given "a" as $t \to \infty$, the diffusion with periodic coefficients $b(\cdot)$, $\beta(\cdot/a)$ is asymptotically Gaussian with the dispersion matrix K given by (24.20), (24.21). For simplicity it will be assumed that the local dispersion matrix is D, where D is a constant positive definite matrix. As before, we assume $X(0) = 0$. Thus

$$X(t) = \int_{[0,t]} b(X(s)) + \beta(X(s)/a)ds + \sigma B(t), \quad D = \sigma\sigma' \tag{24.31}$$

The corresponding diffusion $X(t)$ mod a on the k-dimensional torus $\mathbb{T}_a$ can be shown to converge to equilibrium in variation norm with distance bounded above by $ca^{k/2}\exp\{-c't/a^2\}$ (Bhattacharya (1999); Diaconis and Stroock (1991); Fill (1991)). Unfortunately, this leads to a relaxation time of the order $t >> a^2 \log a$. To get the optimal order $t >> a^2$, we refer to a result of Franke (2004). To state this result, it is convenient to consider

$$Y(t) = X(a^2 t)/a, \quad Z(t) = Y(t) \text{ mod} 1.$$

Now $Z(\cdot)$ is a diffusion on the unit k-dimensional torus $\mathbb{T}_1$, whose generator is

$$A_a = \frac{1}{2} \sum_{1 \le i,j \le k} D_{ij} \frac{\partial^2}{\partial y_i \partial y_j} + (ab(a\cdot) + a\beta(\cdot)) \cdot \nabla. \tag{24.32}$$

It is also the generator of $Y(t)$ but restricted to functions which are periodic with period one. Let $q_{v,D}(t; x, y)$ be the transition probability density of a diffusion on the unit torus with dispersion matrix D as in (24.32) but an arbitrary divergence-free velocity $v(\cdot)$ in place of $ab(a\cdot) + a\beta(\cdot)$. Then the result in Franke (2004) says that

$$\delta := \inf[\min_{x,y \in \mathbb{T}_1} q_{v,D}(1; x, y) : v(\cdot) \text{ is divergence-free}] > 0.$$

It now easily follows from Doeblin's minorization (see BCPT[2] p. 213) that

$$\sup\{\int_{\mathbb{T}_1} |q_{v,D}(t; x, y) - 1| dy : x, y \in \mathbb{T}_1, v(\cdot) \text{ is divergence free}\} \le c' e^{-\delta t},$$

[2] Throughout, BCPT refers to Bhattacharya and Waymire (2016), A Basic Course in Probability Theory.

for some constant $c' > 0$. This immediately leads to

$$\sup_{x \in \mathbb{T}_a} \int \mathbb{T}_a |p_a(t; x, y) - a^{-k}| dy \le c' \exp\{-\delta t/a^2\}, \tag{24.33}$$

where p_a is the transition probability density of $X(t)$ mod$a$, on the big torus $\mathbb{T}_a$. One may now proceed assuming that, over this time scale, $\{X(t) \text{ mod} a : t \ge 0\}$, is an ergodic stationary process. It is convenient to consider the rescaled process

$$Y(t) = X(a^2 t)/a, \quad t \ge 0.$$

Let $Z(t) = Y(t)$ mod1, $t \ge 0$, be a stationary ergodic diffusion on the unit torus $\mathbb{T}_1$. The infinitesimal generator of the diffusion $Z(\cdot)$ on the torus, as well as that of $Y(\cdot)$ restricted to periodic functions, will be denoted by

$$A_a = \frac{1}{2} \sum_{1 \le j \le k} D_{ij} \frac{\partial^2}{\partial y_i \partial y_j} + a(b(a\cdot) + \beta(\cdot)) \cdot \nabla. \tag{24.34}$$

A_a acts on a dense subset $\mathcal{L}(A_a)$ of $L^2(\mathbb{T}_1, dy)$ and, in view of the exponential ergodicity, the range of A_a is $1^\perp$, the set of all mean zero square integrable functions on $\mathbb{T}_1$ (Theorem 17.8). Therefore, there exists a (unique) solution $g_j \in 1^\perp$ of the equation

$$A_a g_j(y) = b_j(ay) + \beta_j(y) - \widetilde{b}_j - \widetilde{\beta}_j (1 \le j \le k), \tag{24.35}$$

where, for an integrable function f,

$$\widetilde{f} = \int_{[0,1]^k} f(y) dy$$

is the mean of f under the invariant distribution $1dy$ on $\mathbb{T}_1$. Writing $g = (g_1, \ldots, g_k)$, and similarly $\widetilde{b}$ and $\widetilde{\beta}$ for the vector means of b and β, it follows by Itô's Lemma that, recalling grad$g(y) := ((\frac{\partial g_i}{\partial y_j}))$,

$$Y(t) - Y(0) - at(\widetilde{b} + \widetilde{\beta}) = a(g(Y(t)) - g(Y(0))) - \int_{[0,t]} [a\text{grad} g(Y(s)) - I_k] \sigma dW(t), \tag{24.36}$$

where $W(t) = B(a^2 t)/a, t \ge 0$, is a standard k-dimensional Brownian motion. It follows from the martingale (functional) central limit theorem[3] that $Y(\cdot)$ is asymptotically a Brownian motion. In particular, for $a/\sqrt{t} \to 0$,

[3] Billingsley (1968), Theorem 23.1. Bhattacharya and Waymire (2022), Theorem 15.5.

$$[Y(t) - Y(0) - at(\widetilde{b} + \widetilde{\beta})]/\sqrt{t} \Rightarrow N(0, K), \tag{24.37}$$

where

$$\begin{aligned} K_{jj'} &= E_{jj'} + D_{jj'}, \\ E_{jj'} &= a^2 \int_{[0,1]^k} \mathrm{grad} g_j(y) D(\mathrm{grad} g_{j'}(y))' dy. \end{aligned} \tag{24.38}$$

This follows from (24.36), and the fact that due to periodic boundary conditions:

$$\int_{[0,1]^k} \frac{\partial g_j}{\partial y'_j} dy = 0, \text{ for all } j, j'. \tag{24.39}$$

The main task now is to analyze the asymptotic behavior of $E_{jj'}$ as a function of "a." We will outline this analysis, leaving some of the details to Bhattacharya (1999). Introduce the complex Hilbert spaces H^0, H^1 as follows:

$H^0 = \{h : \mathbb{R}^k \to \mathbb{C} : h$ is periodic with period lattice $\mathbb{Z}^k$, $\int_{[0,1]^k} |h(y)|^2 dy < \infty$, $\int_{[0,1]^k} h(y)dy = 0\}$, $\quad \langle f, g\rangle_0 = \int_{[0,1]^k} f(y)\overline{g}(y)dy$.
$H^1 = \{h \in H^0 : \int_{[0,1]^k} |\nabla h|^2 (dy) < \infty\}$,

where

$$\begin{aligned} \langle f, g\rangle &= \frac{1}{2} \sum_{1 \le j, j' \le k} \sum_{n \ge 1} \int_{[0,1]^k} \frac{\partial f}{\partial y_j} \frac{\partial \overline{g}}{\partial y'_j} dy \\ &= -\langle \mathscr{D} f, g\rangle_0 \\ &= \frac{1}{2} \int_{[0,1]^k} \mathrm{grad} f(y) D \mathrm{grad} \overline{g}(y) dy, \end{aligned}$$

and $\overline{g}(y)$ denotes the complex conjugate of $g(y)$.

Note that

$$E_{jj'} = a^2 (\langle g_j, g_{j'}\rangle + \langle g_{j'}, g_j\rangle_1). \tag{24.40}$$

It is useful to introduce the second-order differential operator:

$$\mathscr{D} = \frac{1}{2} \sum_{1 \le j, j' \le k} D_{ij} \frac{\partial^2}{\partial y_j \partial y'_j} \tag{24.41}$$

Since $b(ay)$ is rapidly oscillating (with period $1/a$), it may be shown that the term involving $b(ay)$ in (24.35)–(24.36) is well approximated by its average, namely, $\widetilde{b}$, and the generator A_a in (24.36) can be approximated by

$$\overline{A} = \frac{1}{2}\sum_{1\le j\le k} D_{ij}\frac{\partial^2}{\partial y_i \partial y_j} + a(\widetilde{b} + \beta(\cdot)) \cdot \nabla, \quad \widetilde{b} = \int_{[0,1]^k} b(y)dy. \tag{24.42}$$

(see Bhattacharya (1999), pp. 971–976, for details). One may now express A_a and $\overline{A}$ as

$$A_a = \mathscr{D}(I + aS_a), \quad S_a := \mathscr{D}^{-1}(b(a\cdot) + \beta(\cdot)) \cdot \nabla, \tag{24.43}$$

and,

$$\overline{A} = \mathscr{D}(I + a\overline{S}), \quad \overline{S} := \mathscr{D}^{-1}(\widetilde{b} + \beta(\cdot) \cdot \nabla), \tag{24.44}$$

where I denotes the identity operator on H^1. The operators S_a and $\overline{S} : H^0 \to H^1$ are skew symmetric and compact. Skew symmetry follows on integration by parts and the divergence-free condition on $b(\cdot)$ and $\beta(\cdot)$ (Exercise 10). To prove compactness, namely, that any bounded subset of H^0 is mapped into a compact set in H^1, one may use the Fourier transform as an isometry on H^0 (up to a constant multiple) or, equivalently, the Parseval relation.[4] Denote by $\hat{f}(r)$ the Fourier coefficients

$$\hat{f}(r) = \int_{[0,1]^k} f(y)e^{-2\pi i r\cdot y}dy = (1/2\pi)^k \int_{[-\pi,\pi]^k} f(y)e^{-ir\cdot y}dy, \quad (r \in \mathbb{Z}^k).$$

On H^0, $\hat{f}(0) = 0$. Now let H^2 be the subspace of H^1 having square integrable derivatives up to order two. Then $g \to \mathscr{D}g$ is an invertible map on H^2 onto H^0. Indeed, if $f = \mathscr{D}g$, then, on integration by parts,

$$\begin{aligned}\hat{f}(r) &= \int_{[0,1]^k} (\mathscr{D}g)(y)e^{-2\pi i r\cdot y}dy \\ &= (1/2)\sum_{1\le j,j'\le k} D_{jj'}\Big(\int_{[0,1]^k} e^{-2\pi i r\cdot y}\frac{\partial^2 g(y)}{\partial y_i \partial y_j}\Big)dy \\ &= (-2\pi^2)\sum_{1\le j,j'\le k} D_{jj'}(r_j r_{j'})\hat{g}(r),\end{aligned} \tag{24.45}$$

$$\hat{g}(r) = (-1/2\pi^2)\Big(\sum_{1\le j,j'\le k} D_{jj'}(r_j r_{j'})\Big)^{-1}\hat{f}(r), \tag{24.46}$$

$$|\hat{g}(r)| \ge (2\lambda_M \pi^2)^{-1}|\hat{f}(r)|/|r|^2, \quad r \in \mathbb{Z}^k\backslash\{0\}). \tag{24.47}$$

[4] See BCPT, p. 133.

Here λ_M is the maximum eigenvalue of D. We now prove that $\mathscr{D}^{-1} : H^0 \to H^1$ is compact. One needs to show that given a bounded sequence f_n bounded by 1 in norm $|| \cdot ||_0$, there exists a subsequence $f_{n'}$, say, $n' = 1, 2, \ldots$, such that

$$g_{n'} = \mathscr{D}^{-1} f_{n'} \to g := \mathscr{D}^{-1} f,$$

for some $f \in H^0$, the convergence being in H^1. Since a bounded subset of H^0 is weakly compact, by Alaoglu's Theorem (see Lemma 2, Chapter 19), there exists a subsequence $f_{n'}$ converging weakly to some $f \in H^0$. This implies $\hat{f}_{n'}(r) \to \hat{f}(r)$ as $n' \to \infty$, for all $r \in \mathbb{Z}^k \backslash \{0\}$. $\hat{f}_{n'}(0) = 0$, $\hat{f}(0) = 0$. Then (see (24)),

$$\begin{aligned} ||g_{n'} - g||_1^2 &= -\langle \mathscr{D}(g_{n'} - g), g_{n'} - g \rangle_0 \\ &= -\langle f_{n'} - f, \mathscr{D}^{-1}(f_{n'} - f) \rangle_0 \\ &\le (1/2\lambda_m \pi^2)\{ \sum_{0<|r|\le R} |\hat{f}_{n'}(r) - \hat{f}(r)|^2 + (1/R^2) \sum_{|r|>R} |\hat{f}_{n'}(r) - \hat{f}(r)|^2 \} \\ &\le (1/2\lambda_m \pi^2)\{ \sum_{0<|r|\le R} |\hat{f}_{n'}(r) - \hat{f}(r)|^2 + 4/R^2 \}, \end{aligned}$$

since $||f_{n'}||_0 \le 1, ||f||_0 \le 1$. Here λ_m is the smallest eigenvalue of the matrix D. Now, given $\varepsilon > 0$, let $R(\varepsilon)$ be such that $4/R(\varepsilon)^2 < \varepsilon/2$; then choose n' large enough so that the sum in the last expression, with $R = R(\varepsilon)$, is less than $\varepsilon/2$. Hence $||g_{n'} - g||_1^2 < \varepsilon$ for all sufficiently large n. This proves $g_{n'} \to g \in H^1$. Next, note that $(b(a\cdot) + \beta(\cdot)) \cdot \nabla$ and $(\widetilde{b} + \beta(\cdot)) \cdot \nabla$ are bounded operators on H^1 (to H^1) (Exercise 10). Therefore, S_a and $\overline{S}$ are compact.[5] As indicated earlier, one may replace S_a by $\overline{S}$ and A_a by $\overline{A}$, for calculating $E_{jj'}$ for the asymptotic dispersion (24.39). Therefore, instead of using the solutions g_j to (24.35), namely, $A_a g_j = b_j(ay) + \beta_j(y) - \widetilde{b}_j - \widetilde{\beta}_j$, we will consider the solutions to

$$\overline{A} h_j = \widetilde{b}_j + \beta_j(y) - \widetilde{b}_j - \widetilde{\beta}_j = \beta_j(y) - \widetilde{\beta}_j, \ (1 \le j \le k).$$

The latter may be expressed as

$$(I + a\overline{S})h_j = \mathscr{D}^{-1}(\beta_j - \widetilde{\beta}_j), \quad 1 \le j \le k. \tag{24.48}$$

Because S is skew symmetric, $S = iG$, where G is a self adjoint compact operator on H^1. Let $\varphi_n, n = 1, 2, \ldots$, be a complete orthonormal sequence of eigenfunctions of G with corresponding real eigenvalues $\gamma_n, n = 1, 2, \ldots, \gamma_n \to 0$ as $n \to \infty$. The corresponding eigenvalues of $\overline{S}$ are $i\gamma_n, n = 1, \ldots$. One can then have the eigenfunction expansions of h_j, e.g., see Appendix A of Bhattacharya and Waymire (2022)):

[5] See, e.g., Reed and Simon (1972), p. 203.

$$h_j = (h_j)_{\mathcal{N}} + \sum_{n\geq 1} \langle h_j, \varphi_n\rangle_1 \gamma_n,$$

$$\mathscr{D}^{-1}(\beta_j - \widetilde{\beta}) = (\mathscr{D}^{-1}(\beta_j - \widetilde{\beta}))_{\mathcal{N}} + \sum_{n\geq 1} \beta_{j,n}\varphi_n, \tag{24.49}$$

$$\beta_{j,n} := \langle (\mathscr{D}^{-1}(\beta_j - \widetilde{\beta})), \varphi_n\rangle_1,$$

where $(h)_{\mathcal{N}}$ is the projection of h onto the *null space* N of S (or G). The Equation (24.48) then may be expressed in spectral terms as

$$(h_j)_{\mathcal{N}} = (\mathscr{D}^{-1}(\beta_j - \widetilde{\beta}))_N, \tag{24.50}$$

$$\langle h_j, \varphi_n\rangle_1 = \frac{\beta_{j,n}}{1 + ia\gamma_n} \equiv \frac{\langle \mathscr{D}^{-1}(\beta_j - \widetilde{\beta}_j), \varphi_n\rangle_1}{1 + ia\gamma_n}, \; n = 1, 2, \ldots. \tag{24.51}$$

We have arrived at the main result for the final phase.

Theorem 24.3 *Assume A1–A3.*

a. *If $\mathscr{D}^{-1}(\beta_j - \widetilde{\beta})_{\mathcal{N}}$, $j \in J \subset 1, 2, \ldots, k$, are linearly independent elements of H^1, then as $a \to \infty$, for t satisfying*

$$t/a^2 \to \infty, \tag{24.52}$$

one has

$$\frac{1}{a\sqrt{t}}(X_j(t) - X_j(0) - t(\widetilde{b}_j + \widetilde{\beta}), j \in J) \Rightarrow N(0, \Gamma),$$

where the elements of $\Gamma = ((\Gamma_{jj'}))_{j,j'\in J}$ are given by

$$\Gamma_{jj'} = 2\langle \mathscr{D}^{-1}(\beta_j - \widetilde{\beta})_{\mathcal{N}}, \mathscr{D}^{-1}(\beta_{j'} - \widetilde{\beta}_{j'})_{\mathcal{N}}\rangle_1. \tag{24.53}$$

b. *For the case where for a set J of coordinates $\mathscr{D}^{-1}(\beta_j - \widetilde{\beta})_{\mathcal{N}} = 0$, $j \in J$, one has that $\frac{1}{\sqrt{t}}(X_j(t) - X_j(0) - t(\widetilde{b}_j + \widetilde{\beta}) : j \in J)$ is asymptotically Gaussian $N(0, \Gamma)$ if (24.52) holds, where*

$$\Gamma_{jj'} = D_{jj'} + E_{jj'} \quad j, j' \in J, \tag{24.54}$$

where the eigenvalues (and trace) of Γ are bounded away from zero and infinity as $a \to \infty$.

Proof (a) We have shown that $Y(t)$ is asymptotically Gaussian $N(0, K)$ (see (24.37). (24.39)). Recalling $X(0) = Y(0) = 0$ and $Y(t) = X(a^2t)/a$, this says

$$[X(a^2t)/a) - X(0) - at(\widetilde{b} + \widetilde{\beta})]/\sqrt{t}$$

is asymptotically Gaussian $N(0, K)$, as $t \to \infty$. Writing t for a^2t, this statement is equivalent

$$[X(t) - X(0) - t\widetilde{b} + \widetilde{\beta})]/a\sqrt{t}$$

is asymptotically Gaussian $N(0, a^{-2}K)$, if (24.52) holds. Therefore, one needs to prove that $a^{-2}K_{jj'} \to \Gamma_{jj'}$ in (24.53), for $j, j' \in J$, as $a \to \infty$. But this follows from (24.39) with g_j replaced by h_j (see (24.40), (24.49), (24.50)).

In case (b),

$$\mathscr{D}^{-1}(\beta_j - \widetilde{\beta}_j) = \sum_{n \geq 1} \beta_{j,n}\varphi_n,$$

where

$$\beta_{j,n} := \langle \mathscr{D}^{-1}(\beta_j - \widetilde{\beta}), \varphi_n \rangle_1,$$

so that

$$\begin{aligned} E_{jj'} &= a^2\{\langle g_j, g_{j'}\rangle_1 + \langle g_{j'}, gj\rangle_1\} \\ &\sim a^2\{\langle h_j, h_{j'}\rangle_1 + \langle h_{j'}, h_j\rangle_1\} \\ &= a^2 \sum_{n \geq 1} (\beta_{j,n}\beta_{j',n} + \beta_{j',n}\widetilde{\beta}_{j,n})/|1 + ia\gamma_n|^2 \end{aligned}$$

are bounded as $a \to \infty$. ■

Let us now see how Examples 1 and 2 fit into this scheme. In Example 1, let $J = \{1\}$. Here $\widetilde{b} = (c_1, 0)$.The null space $\mathcal{N}$ of $\overline{S}$, or equivalently, of

$$(\widetilde{b} + \beta(\cdot)) \cdot \nabla = \beta(\cdot) \cdot \nabla,$$

comprise functions $g \in H^1$ such that

$$(c_1 + \beta_1(y_1, y_2))\frac{\partial g}{\partial y_1} + \beta_2(y_1, y_2)\frac{\partial g}{\partial y_2} = 0.$$

This implies

$$\frac{\partial g}{\partial y_1} = 0,$$

because $\beta_2 = 0$ and $\beta_1(y) = c_3 \cos 2\pi y_2$. That is, g is a function of y_2 only, $g(y_2)$, say. Thus

$$\mathcal{N} = \{g \in H^1 : g \text{ is a function of } y_2\}.$$

In particular,

$$\mathscr{D}^{-1}(\beta_1 - \widetilde{\beta}_1)_N = \mathscr{D}^{-1}(\beta_1)_N = (-c_3/2\pi^2)\cos 2\pi y_2.$$

Now check that Γ_{11} is as in the last row of Table 24.1. On the other hand, if one takes $J = \{2\}$, then $\beta_2 = 0$, so that $\mathscr{D}^{-1}(\beta_2)_{\mathcal{N}} = 0$, and part (b) applies. It is not necessary here to go through the eigenfunction expansion, directly calculating $\Gamma_{22} = 1$.

In Example 2, let $J = \{1\}$. Then $(\beta_1 - \widetilde{\beta}_1) = c_3 \cos 2\pi y_2$ belongs to the range of

$$(\widetilde{b} + \beta(\cdot)) \cdot \nabla g = (c_1 + c_3 \cos 2\pi y_2)\frac{\partial g}{\partial y_1} + \delta \frac{\partial g}{\partial y_2}, \tag{24.55}$$

i.e., it has no component in the null space, as $(\beta_1 - \widetilde{\beta}_1) = c_3 \cos 2\pi y_2$ belongs to the range of $(\widetilde{b} + \beta(\cdot)) \cdot \nabla g = (c_1 + c_3 \cos 2\pi y_2)\frac{\partial g}{\partial y_1} + \delta \frac{\partial g}{\partial y_2}$ (Exercise 13).

Remark 24.4 The dramatic difference in asymptotic dispersion in the cases (a) and (b) in Theorem 24.3 (and in the examples considered earlier) seems enigmatic. Figures 24.1 and 24.2 from Fried and Combarnous (1971) are indicative of case (a). Case (b) does not seem common in experiments. One may perhaps relate it to the fact that if the generator $\overline{S}$ of a flow $z(t)$ on the torus $\mathbb{T}_1$, namely,

$$dz(t) = a(\widetilde{b} + \beta(\cdot)) \cdot \nabla z(t),$$

has its null space $N = \{0\}$, then the flow is ergodic. In this case the flow is regular and steady.

Remark 24.5 It may be shown (Exercise 14) that the asymptotic dispersion of $X(t)$ defined by

$$X(t) := X(0) + u_0 \int_{[0,t]} \beta(X(s))ds + \sqrt{D}dB(t),$$

as $u_0 \to \infty$, is the same as that of the process $Y(t)$ defined by

$$Y(t) := Y(0) + u_0 \int_{[0,t]} \beta(Y(s))ds + \sqrt{D}dB(t),$$

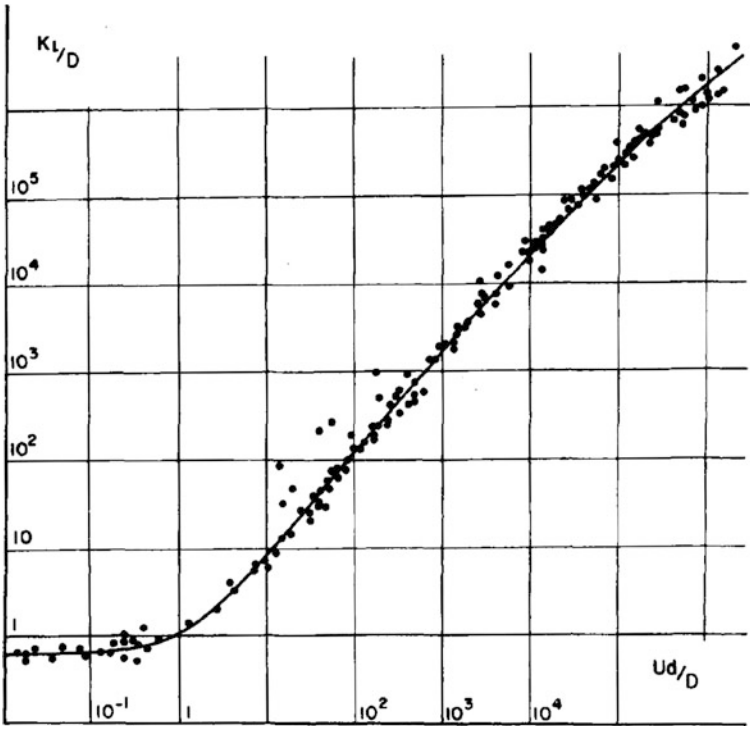


Fig. 24.1 Increased longitudinal dispersion with Peclet number

assuming that β is periodic (see Bhattacharya (1999), pp. 963, 964).

Figure 24.1 below is illustrative of case (a) of Theorem 24.3 as is, of course, Figure 24.2. The Peclet number is a dimensionless number of the generic form dU/D, where d is a characteristic length of the porous media, D is dispersion rate, and U is the average flow velocity, for a specified direction of flow. The Peclet number is a surrogate for the root-mean squared distance $L(t) > 0$ from the mean flow in interpreting experimental results.

The original version of the results in this chapter appeared in Bhattacharya and Goetze (1995). For some additional literature related to this chapter, we refer to Bhattacharya (1985), Bhattacharya et al. (1989), Bhattacharya (1999), Bhattacharya et al. (1999), Avellaneda and Majda (1992), and Owhadi (2003).

Exercises

1. (i) Let a, b, c be positive numbers $c \leq (a + b) \wedge 1$. Show that $c \leq (2 \max\{\sqrt{a}, \sqrt{b}\}) \wedge 1$. [*Hint*: Obvious for $b \geq 1$, or $a \geq 1$. For $a < 1, b < 1$,

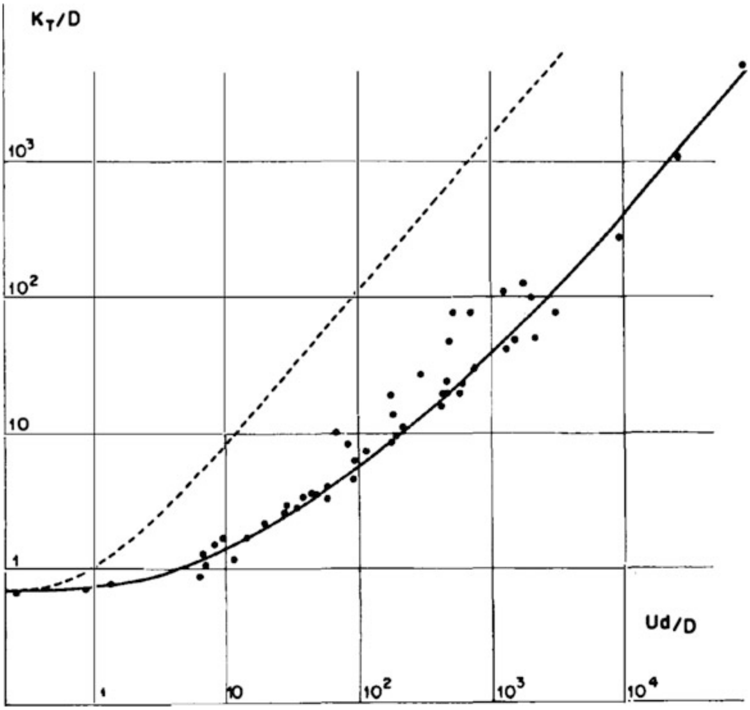


Fig. 24.2 Increased transverse dispersion with Peclet number

$a + b \le \max\{2a, 2b\} < 2\max\{\sqrt{a}, \sqrt{b}.\}$.] (ii) Show that $(c'\gamma + c''\gamma^2) \wedge 1 \le [c' + c'' + 2(c'')^{1/2}]\gamma)$, for nonnegative c', c'', γ.

2. (i) Prove that the invariant distribution π in each of the Examples 1, 2, is the uniform distribution on the torus $[0, 1] \times [0, 1]$ (with 0 and 1 identified)[*Hint*: Compute the adjoint L_0^*, and note that $\mathrm{div}b(\cdot) \equiv \sum_{i=1,2} \frac{\partial b_i}{\partial y_i} = 0$, and $D(\cdot) = D$ is a constant matrix.] (ii) Compute the mean vectors $\widetilde{b}$ and the 2×2 dispersion matrices K of the Gaussian law in Examples 1, 2 directly, without using (24.20).
3. Verify the convergence in the second row of Table 24.1. [*Hint*: The integral of the first term of $X_1(t)$ is of the order t, which, when divided by t^2/a^2, goes to zero in the given range of t. For the second term of $X_1(t)$,Taylor expansion around 0 yields $\cos(2\pi B_2(s)/a) - 1 = -(2\pi^2/a^2)B_2^2(s) + (8\pi^3/6a^2)B_2^2(s)\theta$, where $|\theta| \le 1$, a.s. The second summand on the right in the expansion can be neglected, when its integral over $[0, t]$ is divided by t^2/a^2; for the first summand, use the fact that the processes $\{\sqrt{t}B_2(s/t) : s \ge 0\}$ and $\{B_2(s) : s \ge 0\}$ have the same distribution. The term $B_1(t)/t^2/a^2 \to 0$ in probability, because $t^2/a^2 >> \sqrt{t}$ iff $t >> a^{4/3}$.]
4. Verify the convergence in the third row of Table 24.1. [*Hint*: By Itô's lemma, $\int_{[0,t]} \sin(2\pi B_2(s))ds = -\sin(2\pi B_2(t))/2\pi^2 + (1/\pi)\int_{[0,t]} \cos(2\pi B_2(s))dB_2(s)$;

$\mathbb{E}[\int_{[0,t]} \sin(2\pi B_2(s))ds]^2 = O(t)$, so that $(1/t)\int_{[0,t]} \sin(2\pi B_2(s))ds \to 0$ in probability. Also, $\int_{[0,t]} \cos(2\pi B_2(s)/a)ds =^{\mathcal{L}} \int_{[0,t]} \cos(2\pi B_2(s/a^2))ds = a^2 \int_{[0,1]} \cos(2\pi B_2(u))du$.]

5. Prove that the middle term on the right in (24.28) goes to zero as $a \to \infty$ and $t/a^{2/3} \to r > 0$.[*Hint*: By a Taylor expansion, up to the second derivative, the middle term is seen to be $O(t^{-1/2}\int_{[0,t]}(B_2(s) + \delta s)/a)^2 ds)$, which is of the order of $r^{5/2}a^{-1/3} \to 0$.]
6. Justify the use of Birkhoff's ergodic theorem for the convergence in the last line of (24.27).
7. Show how the martingale CLT can be used to obtain the convergence to the appropriate Gaussian law for the case $t >> a^2$.
8. Instead of the martingale CLT, suggested in Exercise 7, use the time change (Corollary 9.5) to derive the desired convergence for $t >> a^2$.
9. Prove 24.30. [*Hint*: The mean of the quantity on the left goes to zero as $A \to \infty$, and its variance,or square, goes to zero, since the covariance $c(s, s')$, say, of the integrand process goes to zero exponentially fast as |s-s' $\to \infty$, uniformly in "a."]
10. Prove that (i) the operators S_a and $\overline{S}$ are skew symmetric, and (ii) $(b(a\cdot)+\beta(\cdot))\cdot\nabla$, and $(\widetilde{b} + \beta(\cdot)) \cdot \nabla : H^1 \to H^1$, are bounded.
11. Compute the exponent of diffusivity θ; see (24.6) in Tables 24.1, 24.2, in the different ranges of time displayed.
12. Prove that $(\beta_1 - \widetilde{\beta}_1) = c_3 \cos 2\pi y_2$ belongs to the range of $(\widetilde{b} + \beta(\cdot)) \cdot \nabla g = (c_1 + c_3 \cos 2\pi y_2)\frac{\partial g}{\partial y_1} + \delta\frac{\partial g}{\partial y_2}$. [*Hint*: Find $g \in H^1$, of the form $y_1 + dy_2$ mod1 satisfying the equation.]
13. Verify that the approximation of g_j by h_j (see (24.42)) does not affect the asymptotic dispersion in Examples 1, 2.
14. Verify Remark 24.5.

Chapter 25
Special Topic: Skew Random Walk and Skew Brownian Motion

Skew Brownian motion was introduced by Itô and McKean (Illinois J Math 7:181–231, 1963) in a paper devoted to probabilistic constructions of Markov processes satisfying Feller's conditions for the classification of all one-dimensional diffusions. Its eventual foundational role in modeling the presence of certain types of interfacial heterogeneities is now well documented. Applications include solute transport, ecology, finance, and, in general, phenomena associated with the occurrence of sharp heterogeneities. This chapter relies rather heavily on the Skorokhod embedding and Doob's maximal inequality for martingales to provide a skew random walk approximation to skew Brownian motion and may be viewed as an illustration of their use in this context.

For motivation[1] of the processes introduced here, consider a suspension of a dilute concentration $c_0(y)$ of solute in a one-dimensional saturated medium at time zero. Suppose that the dispersion rate is D^- on $(-\infty, 0)$ and D^+ on $(0, \infty)$. If $D^- \neq D^+$ then $y = 0$ is referred to as an *interface*. A basic model in partial differential equation for the time evolution (dispersion) of such a concentration $c(t, y)$ is given by

$$\frac{\partial c}{\partial t} = \frac{1}{2}\frac{\partial}{\partial y}(D(y)\frac{\partial c}{\partial y}), \quad y \neq 0, \quad c(0^+, y) = c_0(y) \tag{25.1}$$

[1] Also see Ouknine (1991), and the survey articles by Lejay (2006), and by Ramirez et al. (2013), as well as the Example 3, Remark 17.10 in Chapter 17. The paper Ramirez (2012) also provides an intriguing application to population biology.

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_25

together with a condition of *continuity of flux* across the interface

$$D^+ \frac{\partial c}{\partial y}(t, 0^+) - D^- \frac{\partial c}{\partial y}(t, 0^-) = 0. \tag{25.2}$$

In the case $D^+ = D^- = D$, as explained in Chapter 2, Example 4, the solution may be represented as

$$c(t, y) = \mathbb{E}_y c_0(Y_t) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi Dt}} e^{-\frac{(y-z)^2}{2Dt}} c_0(z) dz, \quad t \geq 0, \tag{25.3}$$

where Y is a Brownian motion with zero drift and diffusion coefficient D (Exercise 6). It is natural to inquire about the nature of the particle motion in the presence of an interface. The answer will follow from the considerations to follow.

The *skew Brownian motion with parameter* $\alpha \in (0, 1)$, $B^{(\alpha)}$ is a one-dimensional stochastic process on $(-\infty, \infty)$ that behaves like a standard Brownian motion on each of the half-lines $(-\infty, 0)$ and $(0, \infty)$. In particular, we will see that $B^{(\alpha)}$ is a strong Markov process having continuous sample paths and homogeneous transition probability density $p^{(\alpha)}(t; x, y)$ satisfying

$$\frac{\partial p^{(\alpha)}}{\partial t} = \frac{1}{2} \frac{\partial^2 p^{(\alpha)}}{\partial x^2}, \quad x, y \in \mathbb{R}, t > 0. \tag{25.4}$$

However, at the interface $x = 0$, one has the following relation:

$$\alpha \frac{\partial p^{(\alpha)}(t; x, y)}{\partial x}|_{x=0^+} - (1 - \alpha) \frac{\partial p^{(\alpha)}(t; x, y)}{\partial x}|_{x=0^-} = 0, \quad t > 0, -\infty < y < \infty. \tag{25.5}$$

In the case $\alpha = \frac{1}{2}$, the interface condition is simply continuity of the derivative at the backward variable $x = 0$, and the transition probabilities therefore coincide with those of a standard Brownian motion.

To cast this in the framework of Feller's semigroup theory, define speed and scale functions by

$$s(x) = \begin{cases} \alpha^{-1} x, & x \geq 0, \\ (1 - \alpha)^{-1} x, & x < 0, \end{cases} \qquad m(x) = \begin{cases} 2\alpha x, & x \geq 0, \\ 2(1 - \alpha) x. & x < 0, \end{cases} \quad . \tag{25.6}$$

Then one has

$$L = D_m D_s^+ = \frac{1}{2} \frac{d^2}{dx^2} \text{ on } \mathbb{R} \backslash \{0\}, \tag{25.7}$$

and

$$D_s f(0^+) - D_s f(0^-) = \frac{df}{ds}(x)|_{x=0^+} - \frac{df}{ds}(x)|_{x=0^-} = \alpha f'(0^+) - (1-\alpha) f'(0^-). \tag{25.8}$$

Thus, in view of (21.5), $f \in \mathcal{D}_L$ if and only if f'' exists and is continuous on $(-\infty, 0) \cup (0, \infty)$, $f''(0^+) = f''(0^-)$, and $\alpha f'(0^+) = (1-\alpha) f'(0^-)$. The inaccessibility of the boundaries $\pm\infty$ imposes no further restrictions on the domain; see Definition 21.3. However, it is also possible to give a direct probabilistic construction of α-skew Brownian motion from standard Brownian motion B. Recall that by continuity of the paths of the Brownian motion, the (random) set $\{t : B_t \neq 0\}$ is the complement of the inverse image of the closed set $\{0\}$ and, therefore, an open subset of $(0, \infty)$. Using the fact that every open subset of $\mathbb{R}$ is a countable disjoint union of open intervals, one may express $\{t : B_t \neq 0\}$ as a countable disjoint union of open intervals $J_1^o, J_2^o, \ldots$, referred to as *excursion intervals* of B. So, while the excursions of Brownian motion cannot be ordered, like the rational numbers they can be enumerated. Note that $B_t = 0$ at an endpoint t of any J_k^o. We denote the closure of J_k^o by J_k. Let $A_1, A_2, \ldots$ be an i.i.d. sequence of ± 1 Bernoulli random variables, independent of B, with $P(A_n = 1) = \alpha$. The α-skew Brownian motion starting at zero is the process $B^{(\alpha)}$ defined by

$$B_t^{(\alpha)} = \sum_{k=1}^{\infty} A_k \mathbf{1}_{J_k}(t)|B_t|, \quad t \geq 0. \tag{25.9}$$

That is, one makes independent tosses of a coin to decide if a particular excursion of $|B|$ should remain positive or should undergo a sign change.

Proposition 25.1 *$B^{(\alpha)}$ is a Markov process having continuous sample paths and stationary transition probabilities $p^{(\alpha)}(t; x, y)$ given by*

$$p^{(\alpha)}(t; x, y) = \begin{cases} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}} + \frac{(2\alpha-1)}{\sqrt{2\pi t}} e^{-\frac{(y+x)^2}{2t}} & \text{if } x \geq 0, y > 0 \\[2ex] \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}} - \frac{(2\alpha-1)}{\sqrt{2\pi t}} e^{-\frac{(y+x)^2}{2t}} & \text{if } x \leq 0, y < 0 \\[2ex] \frac{2\alpha}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}} \quad \text{if } x \leq 0, y > 0 \\[2ex] \frac{2(1-\alpha)}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}} \quad \text{if } x \geq 0, y < 0. \end{cases}$$

Proof Let $k(t) = k$ for $t \in J_k^o$, and take the right-continuous extension to extend to the value at an endpoint t; i.e., $k(t) = k$ when t is the left endpoint of J_k. Then $B_t^{(\alpha)} = A_{k(t)}|B_t|$. Also $k(t)$ is constant on an open interval. It follows that $t \to B_t^{(\alpha)}$ is continuous on each excursion interval. Moreover, since $B_t = 0$ for $t \in \partial J_k$, $k \geq 1$, the continuity extends to the endpoints of excursion intervals. Let $\mathcal{F}_t := \sigma\{|B_s|, 0 \leq$

$s \le t\} \vee \sigma\{A_1, A_2, \dots\} \supset \sigma(B_s^{(\alpha)}, s \le t)$. Then, for $0 \le s < t$ and a nonnegative, measurable function g one has using the Markov property for $|B|$ and independence of $|B|$ of the i.i.d. sign changes $A_1, A_2, \dots,$

$$
\begin{aligned}
&\mathbb{E}\{g(B_t^{(\alpha)})|\mathcal{F}_s\} \\
&\quad = \mathbb{E}\{g(A_{k(t)}|B_t|)|\mathcal{F}_s\} \\
&\quad = \sum_{k=1}^{\infty} \mathbb{E}\{g(A_k|B_t|)\mathbf{1}_{J_k}(t)|\mathcal{F}_s\} \\
&\quad = \mathbf{1}[A_{k(t)} = 1] \sum_{k=1}^{\infty} \mathbb{E}\{g(|B_t|)\mathbf{1}_{J_k}(t)|\mathcal{F}_s\} \\
&\qquad + \mathbf{1}[A_{k(t)} = -1] \sum_{k=1}^{\infty} \mathbb{E}\{g(-|B_t|)\mathbf{1}_{J_k}(t)|\mathcal{F}_s\} \\
&\quad = \mathbf{1}[A_{k(t)} = 1]\mathbb{E}\{g(|B_t|)|\sigma(|B_s|, A_k, k \ge 1)\} \\
&\qquad + \mathbf{1}[A_{k(t)} = -1]\mathbb{E}\{g(-|B_t|)|\sigma(|B_s|, A_k, k \ge 1)\} \\
&\quad = \mathbb{E}\{g(A_{k(t)}|B_t|)|\sigma(|B_s|, A_k, k \ge 1)\} \\
&\quad = \mathbb{E}\{g(B_t^{(\alpha)})|\sigma(|B_s|, A_k, k \ge 1)\} = \mathbb{E}\{g(B_t^{(\alpha)})|\sigma(B_s^{(\alpha))}\}.
\end{aligned}
$$

The Markov property follows from the smoothing property of conditional expectation (see BCPT[2] p. 38), since $\mathcal{F}_s \supset \sigma(B_u^{(\alpha)}, u \le s)$. The computation of the transition probabilities $p^{(\alpha)}(t; x, y)$ will follow from the strong Markov property for Brownian motion: specifically, the reflection principle. The bulk of these are left as Exercise 3; however observe, for example, that for $y > 0$,

$$
\begin{aligned}
&P_0(B_t^{(\alpha)} > y) \\
&\quad = \sum_{n=1}^{\infty} [P_0(B_t > y, A_n = +1, t \in J_n) + P_0(B_t < -y, A_n = +1, t \in J_n)] \\
&\quad = \alpha P_0(B_t > y) + \alpha P_0(B_t < -y).
\end{aligned} \tag{25.10}
$$

In particular, it follows that

$$
p^{(\alpha)}(t; 0, y) = \frac{2\alpha}{\sqrt{2\pi t}} e^{-\frac{y^2}{2t}}.
$$

■

[2] Throughout, BCPT refers to Bhattacharya and Waymire (2016), A Basic Course in Probability Theory.

One may check that for fixed y, the transition probabilities of the skew Brownian motion are continuous in the backward variable x. This implies the Feller property (see Chapter 1, Definition 1.1). The Feller property enjoys numerous important consequences in the general theory of Markov processes. The following result is chief among these when taken together with the sample path continuity.

Theorem 25.2 *Skew Brownian motion has the strong Markov property.*

Proof In view of the sample path continuity and the Feller property, the strong Markov property follows by Theorem 1.6. ■

The stochastic process corresponding to the interfacial dispersion problem posed at the outset of this chapter now follows by a rescaling of an α-skew Brownian motion for an appropriate choice of α. Specifically one has the following consequence; see Exercise 9.

Corollary 25.3 *The process defined by $Y = s(B^{(\alpha^*)})$ where $s(y) = \sqrt{D^+}y$, $y \geq 0$, and $s(y) = \sqrt{D^-}y$, $y \leq 0$, and $\alpha^* = \frac{\sqrt{D^+}}{\sqrt{D^+}+\sqrt{D^-}}$ is a Markov process having continuous sample paths with stationary transition probabilities given by*

$$
\begin{aligned}
&p(t; x, y) \\
&= \begin{cases}
\frac{1}{\sqrt{2\pi D^+ t}}[\exp\{-\frac{(y-x)^2}{2D^+t}\} + \frac{\sqrt{D^+}-\sqrt{D^-}}{\sqrt{D^-}+\sqrt{D^+}}\exp\{-\frac{(y+x)^2}{2D^+t}\}] & x > 0, y > 0, \\
\frac{1}{\sqrt{2\pi D^- t}}[\exp\{-\frac{(y-x)^2}{2D^-t}\} - \frac{\sqrt{D^+}-\sqrt{D^-}}{\sqrt{D^+}+\sqrt{D^-}}\exp\{-\frac{(y+x)^2}{2D^-t}\}] & x < 0, y < 0, \\
\frac{2}{\sqrt{D^+}+\sqrt{D^-}}\frac{1}{\sqrt{2\pi t}}\exp\{-\frac{(y\sqrt{D^-}-x\sqrt{D^+})^2}{2D^-D^+t}\} & x \leq 0, y > 0, \\
\frac{2}{\sqrt{D^+}+\sqrt{D^-}}\frac{1}{\sqrt{2\pi t}}\exp\{-\frac{(y\sqrt{D^+}-x\sqrt{D^-})^2}{2D^-D^+t}\} & x \geq 0, y < 0.
\end{cases}
\end{aligned}
\tag{25.11}
$$

Moreover, for each fixed $x \in \mathbb{R}$, one has

$$\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial}{\partial y}(D(y)\frac{\partial p(t; x, y)}{\partial y}), \quad y \neq 0, \tag{25.12}$$

together with continuity of flux across the interface

$$D^+\frac{\partial p}{\partial y}(t; x, 0^+) - D^-\frac{\partial p}{\partial y}(t; x, 0^-) = 0. \tag{25.13}$$

Note that, unlike the case of skew Brownian motion noted earlier, (25.11) is jointly continuous in the forward and backward transition variables x, y. In fact $p(t; x, y)$ is symmetric in x, y. However, this symmetry is not a property of the transition probabilities for skew Brownian motion given in Proposition 25.1.

Skew random walk is a natural discretization of skew Brownian motion defined as follows.

Definition 25.1 The *skew random walk* is a discrete Markov chain $\{Y_n : n = 0, 1, 2, \dots\}$ on the integers $\mathbb{Z}$ having transition probabilities

$$p_{ij}^{(\alpha)} = \begin{cases} \frac{1}{2} \ \text{ if } i \neq 0,\ j = i \pm 1 \\ \\ \alpha \ \text{ if } i = 0,\ j = 1 \\ \\ 1-\alpha \ \text{ if } i = 0,\ j = -1. \end{cases}$$

Remark 25.1 One may note that the Markov process does not *skip states* and, as such, is an example of a discrete parameter birth–death Markov chain. In fact, the terminology "skew random walk"is a misnomer since the process does *not* have independent increments.[3]

One may anticipate an approximation like that of the functional central limit theorem connecting random walks to Brownian motion will also hold here.[4] However, in view of the non-Lipschitz nature of the coefficients, the martingale approximation developed in Chapter 23, Example 5 does not directly apply here. The remainder of this chapter gives a proof based on the Skorokhod embedding method for skew Brownian motion.[5]

Let us recall that by an application of the Skorokhod embedding theorem,[6] there is a sequence of times $T_1 < T_2 < \cdots$ such that B_{T_1} has a symmetric Bernoulli ± 1-distribution, and $B_{T_{i+1}} - B_{T_i}$ $(i \geq 0)$ are i.i.d. with with a symmetric ± 1-distribution. Moreover, $T_{i+1} - T_i$ $(i \geq 0)$ are i.i.d. with mean one; also see Exercise 5.

Lemma 1 (Discrete Excursion Representation) *Let $S = \{S_n : n = 0, 1, 2, \dots\}$ be a simple symmetric random walk starting at 0, and let $\tilde{J}_{\pi_1}, \tilde{J}_{\pi_2}, \dots$ denote an enumeration of the excursions of S away from zero for a fixed but arbitrary permutation π of the natural numbers. In particular $|S_n| > 0$ if $n \in \tilde{J}_{\pi_k}$. Define*

[3] Iksanov and Pilipenko (2023) and Dong et al. (2023) recently extended the theory to skew stable Lévy process as a limit of a sequence of stable Lévy processes perturbed at zero.

[4] Convergence of the distribution at a fixed time point was first announced in Harrison and Shepp (1981). The suggested "fourth moment proof"along the lines of that given for simple symmetric random walk (i.e., $\alpha = 1/2$) based on convergence of finite dimensional distributions and tightness is quite laborious and tricky due to the lack of independence of the increments. A full proof was given by Brooks and Chacon (1983).

[5] Sooie-Ho Loke co-authored this approach while a graduate student at Oregon State University. However, a more general functional central limit theorem by Cherny et al. (2002) appeared earlier, so that the collaboration with Loke remained unpublished until this book chapter.

[6] See BCPT, p. 200, or Bhattacharya and Waymire (2021), p.204.

$$S_0^{(\alpha)} = 0, \quad S_n^{(\alpha)} = \sum_{k=1}^{\infty} \mathbf{1}_{\tilde{J}_{\pi_k}}(n)\tilde{A}_k|S_n|, n \geq 1,$$

where $\tilde{A}_1, \tilde{A}_2, \dots$ *is an i.i.d. sequence of Bernoulli* ± 1*-random variables, independent of* S*, with* $P(\tilde{A}_1 = 1) = \alpha$*. Then* $S^{(\alpha)}$ *is distributed as* α*-skew random walk.*

Proof Let G be a bounded $\sigma(S_j^{(\alpha)}, j \leq n)$ measurable function. We will first show that $P(S_{n+1}^{(\alpha)} = S_n^{(\alpha)} + 1 \mid S_j^{(\alpha)}, j \leq n) = 1/2$ on $[S_n^{(\alpha)} > 0]$. For this consider

$$\begin{aligned}
&\mathbb{E}\{G(S_1^{(\alpha)}, \dots, S_n^{(\alpha)})\mathbf{1}_{[S_{n+1}^{(\alpha)} = S_n^{(\alpha)}+1]}\mathbf{1}_{[S_n^{(\alpha)}>0]}\} \\
&\quad = \mathbb{E}\{\mathbb{E}[G(S_1^{(\alpha)}, \dots, S_n^{(\alpha)})\mathbf{1}_{[S_{n+1}^{(\alpha)} = S_n^{(\alpha)}+1]}\mathbf{1}_{[S_n^{(\alpha)}>0]} \mid S_j^{(\alpha)}, j \leq n]\} \\
&\quad = \mathbb{E}\{G(S_1^{(\alpha)}, \dots, S_n^{(\alpha)})\mathbb{E}[\mathbf{1}_{[S_{n+1}^{(\alpha)} = S_n^{(\alpha)}+1]}\mathbf{1}_{[S_n^{(\alpha)}>0]} \mid S_j^{(\alpha)}, j \leq n]\} \\
&\quad = \mathbb{E}\{G(S_1^{(\alpha)}, \dots, S_n^{(\alpha)})\mathbb{E}[\mathbf{1}_{[S_{n+1}^{(\alpha)} = S_n^{(\alpha)}+1]}\mathbf{1}_{[S_n^{(\alpha)}>0]}\mathbf{1}_{[S_{n+1}^{(\alpha)}>0]} \mid S_j^{(\alpha)}, j \leq n]\} \\
&\quad = \mathbb{E}\{G(S_1^{(\alpha)}, \dots, S_n^{(\alpha)})\mathbb{E}[\mathbf{1}_{[|S_{n+1}| = |S_n|+1]}\mathbf{1}_{[S_n^{(\alpha)}>0]}\mathbf{1}_{[S_{n+1}^{(\alpha)}>0]} \mid S_j^{(\alpha)}, j \leq n]\} \\
&\quad = \mathbb{E}\{G(S_1^{(\alpha)}, \dots, S_n^{(\alpha)})\mathbb{E}[\mathbf{1}_{[|S_{n+1}| = |S_n|+1]}\mathbf{1}_{[S_n^{(\alpha)}>0]} \mid S_j^{(\alpha)}, j \leq n]\} \\
&\quad = \mathbb{E}\{G(S_1^{(\alpha)}, \dots, S_n^{(\alpha)})\mathbf{1}_{[S_n^{(\alpha)}>0]}\mathbb{E}[\mathbf{1}_{[|S_{n+1}| = |S_n|+1]} \mid S_j^{(\alpha)}, j \leq n]\} \\
&\quad = \mathbb{E}\left\{G(S_1^{(\alpha)}, \dots, S_n^{(\alpha)})\mathbf{1}_{[S_n^{(\alpha)}>0]} \cdot \frac{1}{2}\right\}.
\end{aligned}$$

The proof that $P(S_{n+1}^{(\alpha)} = S_n^{(\alpha)} + 1 \mid S_j^{(\alpha)}, j \leq n) = 1/2$ on $[S_n^{(\alpha)} < 0]$ is by entirely similar considerations and left to the reader.

So let us turn to showing that $P(S_{n+1}^{(\alpha)} = S_n^{(\alpha)} + 1 \mid S_j^{(\alpha)}, j \leq n) = \alpha$ on $[S_n^{(\alpha)} = 0]$. For this consider

$$\begin{aligned}
&\mathbb{E}\{G(S_1^{(\alpha)}, \dots, S_n^{(\alpha)})\mathbf{1}_{[S_{n+1}^{(\alpha)} = S_n^{(\alpha)}+1]}\mathbf{1}_{[S_n^{(\alpha)}=0]}\} \\
&\quad = \mathbb{E}\{\mathbb{E}[G(S_1^{(\alpha)}, \dots, S_n^{(\alpha)})\mathbf{1}_{[S_{n+1}^{(\alpha)} = 1]}\mathbf{1}_{[S_n^{(\alpha)}=0]} \mid \{S_j^{(\alpha)}\}_1^n, S_n^{(\alpha)} = 0]\} \\
&\quad = \mathbb{E}\{\mathbb{E}[G(S_1^{(\alpha)}, \dots, S_n^{(\alpha)})\mathbf{1}_{[n+1 \in J_{\pi(S_1^{(\alpha)}, \dots, S_n^{(\alpha)})}]}\mathbf{1}_{[A_{\pi(S_1^{(\alpha)}, \dots, S_n^{(\alpha)})}=1]} \\
&\qquad \times \mathbf{1}_{[|S_{n+1}|=1]}\mathbf{1}_{[S_n^{(\alpha)}=0]} \mid \{S_j^{(\alpha)}\}_1^n, S_n^{(\alpha)} = 0]\} \\
&\quad = \mathbb{E}\{\mathbb{E}[G(S_1^{(\alpha)}, \dots, S_n^{(\alpha)})\mathbf{1}_{[n+1 \in J_{\pi(S_1^{(\alpha)}, \dots, S_n^{(\alpha)})}]}\mathbf{1}_{[A_{\pi(S_1^{(\alpha)}, \dots, S_n^{(\alpha)})}=1]} \\
&\qquad \times \mathbf{1}_{[|S_n|=0]}\mathbf{1}_{[S_n^{(\alpha)}=0]} \mid \{S_j^{(\alpha)}\}_1^n, S_n^{(\alpha)} = 0]\}
\end{aligned}$$

$$= \mathbb{E}\{G(S_1^{(\alpha)}, \ldots, S_n^{(\alpha)})\mathbf{1}_{[S_n^{(\alpha)}=0]}\mathbf{1}_{[n+1\in J_{\pi(S_1^{(\alpha)},\ldots,S_n^{(\alpha)})}]}$$

$$\times\mathbb{E}[\mathbf{1}_{[A_{\pi(S_1^{(\alpha)},\ldots,S_n^{(\alpha)})}=1]}\mathbf{1}_{[|S_{n+1}|=1]}\mathbf{1}_{[|S_n|=0]} \mid \{S_j^{(\alpha)}\}_1^n, S_n^{(\alpha)} = 0]\}$$

Now since the event $[|S_n| = 0]$ implies the event $[|S_{n+1}| = 1]$, the indicator random variable $\mathbf{1}_{[|S_{n+1}|=1]}\mathbf{1}_{[|S_n|=0]}$ is measurable with respect to the σ-field $\sigma(\{S_j^{(\alpha)}\}_1^n, S_n^{(\alpha)} = 0)$. Also since the event $[|S_n| = 0]$ implies that $|S_{n+1}|$ is in a completely new excursion and the fact that $\{A_k\}$ is independent of $\{S_k\}$, $\mathbf{1}_{[A_{\pi(S_1^{(\alpha)},\ldots,S_n^{(\alpha)})}=1]}$ is independent of $\sigma(\{S_j^{(\alpha)}\}_1^n, S_n^{(\alpha)} = 0)$. Thus we have,

$$\mathbb{E}[G(S_1^{(\alpha)}, \ldots, S_n^{(\alpha)})\mathbf{1}_{[S_{n+1}^{(\alpha)}=S_n^{(\alpha)}+1]}\mathbf{1}_{[S_n^{(\alpha)}=0]}]$$

$$= \mathbb{E}[G(S_1^{(\alpha)}, \ldots, S_n^{(\alpha)})\mathbf{1}_{[S_n^{(\alpha)}=0]}\mathbf{1}_{[n+1\in J_{\pi(S_1^{(\alpha)},\ldots,S_n^{(\alpha)})}]}$$

$$\times \mathbb{E}[\mathbf{1}_{[A_{\pi(S_1^{(\alpha)},\ldots,S_n^{(\alpha)})}=1]}]\mathbb{E}[\mathbf{1}_{[|S_{n+1}|=1]}\mathbf{1}_{[|S_n|=0]} \mid \{S_j^{(\alpha)}\}_1^n, S_n^{(\alpha)} = 0]]$$

$$= \mathbb{E}[G(S_1^{(\alpha)}, \ldots, S_n^{(\alpha)})\mathbf{1}_{[S_n^{(\alpha)}=0]} \cdot \alpha]$$

■

The proof of the next lemma is left as an exercise.

Lemma 2 *The discrete parameter stochastic process* $\widetilde{S}_0^{(\alpha)} = 0, \widetilde{S}_m^{(\alpha)} = B_{T_m}^{(\alpha)}, m = 1, 2, \ldots,$ *is distributed as an* α*-skew random walk.*

Lemma 3 *For any* $\lambda > 0$*, the distribution of* α*-skew Brownian motion is invariant under the space-time transformation* $\{\lambda^{-\frac{1}{2}} B_{\lambda t}^{(\alpha)} : t \geq 0\}$.

Proof Note that the Markov property and continuity of sample paths are preserved by this continuous bijective transformation on path space $C[0, \infty)$. So it is sufficient to check, by inspection, that

$$\lambda^{\frac{1}{2}} p^{(\alpha)}(\lambda t; \lambda^{\frac{1}{2}} x, \lambda^{\frac{1}{2}} y) = p^{(\alpha)}(t; x, y), \quad x, y \in \mathbb{R}, t \geq 0.$$

■

The following lemma might be viewed as a symmetrization reminiscent of *von Neumann's algorithm* for simulating a fair coin by tossing it twice;[7] also see Exercise 4.

Lemma 4 (Walsh's Martingale) *Define a reversal function*

[7] This martingale was identified by Walsh (1978), who also pointed out skew Brownian motion as an example of a (continuous) diffusion having a discontinuous local time; also see Ramirez et al. (2006) in this regard.

$$r_\alpha(x) = \alpha x \mathbf{1}_{(-\infty,0]}(x) + (1-\alpha)x\mathbf{1}_{[0,\infty)}(x),\, x \in \mathbb{R}.$$

Then $r_\alpha(B^{(\alpha)})$ is a martingale.

Proof Fix α and let $R_t = r_\alpha(B_t^{(\alpha)}), t \geq 0$. A proof can be obtained by a direct calculation of $\mathbb{E}\{R_t|\sigma(B_u : u \leq s)\}$ via the Markov property and formulae for the transition probabilities of the skew Brownian motion. First consider the case $B_s^{(\alpha)} > 0$. Then, as a result of the indicated cancellations of integrals, one has

$$\begin{aligned}
&\mathbb{E}\{r_\alpha(B_t^{(\alpha)})|\sigma(B_u : u \leq s)\} \\
&= \int_{\mathbb{R}} r_\alpha(y)p^{(\alpha)}(t-s;\, B_s^{(\alpha)}, y)dy \\
&= (1-\alpha)\int_0^\infty yp^{(\alpha)}(t-s;\, B_s^{(\alpha)}, y)dy - \alpha\int_0^\infty yp^{(\alpha)}(t-s;\, B_s^{(\alpha)}, -y)dy \\
&= \frac{(1-\alpha)}{\sqrt{2\pi(t-s)}}\int_0^\infty y\{e^{-\frac{(y-B_s^{(\alpha)})^2}{2(t-s)}} - e^{-\frac{(y+B_s^{(\alpha)})^2}{2(t-s)}}\}dy \\
&= r_\alpha(B_s^{(\alpha)}).
\end{aligned}$$

The last equality follows by observing that the last integral is the mean of a Brownian motion starting at $B_s^{(\alpha)}$ and absorbed at zero. The case that $B_s^{(\alpha)} \leq 0$ is similar and left to the reader. ■

Lemma 5 *For each $k = 1, 2, \ldots,$ and $m \geq 2$,*

$$\mathbb{E}|r_\alpha(B_{k+1}^{(\alpha)}) - r_\alpha(B_k^{(\alpha)})|^m \leq 1.$$

Proof Similarly to the previous lemma, a proof can be obtained by a direct calculation via the Markov property and formulae for the transition probabilities of the skew Brownian motion. Namely,

$$\begin{aligned}
&\mathbb{E}|r_\alpha(B_{k+1}^{(\alpha)}) - r_\alpha(B_k^{(\alpha)})|^m \\
&= (1-\alpha)^m\int_0^\infty\int_0^\infty |y-x|^m p(1;\, x, y)p(k;\, 0, x)dydx \\
&\quad +\alpha^m\int_{-\infty}^0\int_{-\infty}^0 |y-x|^m p(1;\, x, y)p(k;\, 0, x)dydx \\
&\quad +\int_0^\infty\int_{-\infty}^0 |\alpha y-(1-\alpha)x|^m p(1;\, x, y)p(k;\, 0, x)dydx \\
&\quad +\int_{-\infty}^0\int_0^\infty |(1-\alpha)y-\alpha x|^m p(1;\, x, y)p(k;\, 0, x)dydx.
\end{aligned}$$

For each of these terms, substituting the formulae for the transition probabilities $p(1; x, y)$, one may check that each of the inner integrals (with respect to y) is bounded by one for any x and α. Adding up the integrals with respect to x yields an upper bound by a total probability of one. ∎

The proofs of the following properties of the reversal function are straightforward to check and left to the reader.

Lemma 6 *For any nonnegative function f on $\mathbb{R}$*

(i) $r_\alpha(\max_x f(x)) = \max_x r_\alpha(f(x))$.
(ii) *For any* $x \in \mathbb{R}$, $r_\alpha(|x|) \leq |r_\alpha(x)|$ iff $\alpha \geq 1/2$.
(iii) *For* $x \geq y$, $|r_\alpha(x - y)| \leq |r_\alpha(x) - r_\alpha(y)|$ iff $\alpha \geq 1/2$.

Define the polygonal random function $S^{(\alpha,n)}$ on $[0, 1]$ as follows:

$$S_t^{(\alpha,n)} := \frac{S_{k-1}^{(\alpha)}}{\sqrt{n}} + n\left(t - \frac{k-1}{n}\right)\frac{S_k^{(\alpha)} - S_{k-1}^{(\alpha)}}{\sqrt{n}}$$
$$\text{for } t \in \left[\tfrac{k-1}{n}, \tfrac{k}{n}\right],\ 1 \leq k \leq n. \tag{25.14}$$

That is, $S_t^{(\alpha,n)} = \frac{S_k^{(\alpha)}}{\sqrt{n}}$ at points $t = \frac{k}{n}$ $(0 \leq k \leq n)$, and $t \mapsto S_t^{(\alpha,n)}$ is linearly interpolated between the endpoints of each interval $[\frac{k-1}{n}, \frac{k}{n}]$.

Theorem 25.4 *$S^{(\alpha,n)}$ converges in distribution to the α-skew Brownian motion $B^{(\alpha)}$ as $n \to \infty$.*

Proof Let T_k, $k \geq 1$, be the embedding times of simple symmetric random walk $\{S_k = B_{T_k} : k = 0, 1, 2, \ldots\}$ in a standard Brownian motion $\{B_t : t \geq 0\}$ starting at zero. That is, the simple symmetric random walk $\{S_k : k = 0, 1, 2, \ldots\}$ has the same distribution as $\{\widetilde{S}_k := B_{T_k} : k = 0, 1, 2, \ldots\}$. It now follows that the skew random walk $\{S_m^{(\alpha)} : m = 0, 1, 2, \ldots\}$ has the same distribution as $\{\widetilde{S}_m^{(\alpha)} \equiv B_{T_m}^{(\alpha)} : m = 0, 1, 2, \ldots\}$. In particular, therefore, $S^{(\alpha,n)}$ has the same distribution as $\widetilde{S}^{(\alpha,n)}$ defined by $\widetilde{S}_{k/n}^{(\alpha,n)} := n^{-\frac{1}{2}} B_{T_k}^{(\alpha)}$ $(k = 0, 1, \ldots, n)$ and with linear interpolation between k/n and $(k+1)/n$ $(k = 0, 1, \ldots, n-1)$. Also, define, for each $n = 1, 2, \ldots$, the skew Brownian motion $\widetilde{B}_t^{(\alpha,n)} := n^{-\frac{1}{2}} B_{nt}^{(\alpha)}$, $t \geq 0$. We will show that

$$\max_{0 \leq t \leq 1} |\widetilde{S}_t^{(\alpha,n)} - \widetilde{B}_t^{(\alpha,n)}| \longrightarrow 0 \quad \text{in probability as } n \to \infty, \tag{25.15}$$

which implies the asserted weak convergence (see Exercise 2). Now

$$\max_{0 \leq t \leq 1} |\widetilde{S}_t^{(\alpha,n)} - \widetilde{B}_t^{(\alpha,n)}| \tag{25.16}$$

$$\leq n^{-\frac{1}{2}} \max_{1\leq k\leq n} |B_{T_k}^{(\alpha)} - B_k^{(\alpha)}|$$

$$+ \max_{0\leq k\leq n-1} \{ \max_{\frac{k}{n}\leq t\leq \frac{k+1}{n}} |\widetilde{S}_t^{(\alpha,n)} - \widetilde{S}_{k/n}^{(\alpha,n)}| + n^{-\frac{1}{2}} \max_{k\leq t\leq k+1} |B_t^{(\alpha)} - B_k^{(\alpha)}|\}$$

$$= I_n^{(1)} + I_n^{(2)} + I_n^{(3)}, \quad \text{say.} \tag{25.17}$$

Now, writing $\tilde{Z}_k = \tilde{S^{(\alpha)}}_k - \tilde{S^{(\alpha)}}_{k-1}$, as $n \to \infty$ one readily has

$$I_n^{(2)} \leq n^{-\frac{1}{2}} \max\{|\tilde{Z}_k| : 1 \leq k \leq n\} \leq \frac{1}{\sqrt{n}} \to 0.$$

For the third term, let $\varepsilon > 0$ and fix a $k \in \{0, 1, \ldots, n-1\}$. Note that using the lemma on properties of the reversal function one has

$$\begin{aligned}
&P(\max_{k\leq t\leq k+1} |B_t^{(\alpha)} - B_k^{(\alpha)}| > \sqrt{n}\varepsilon) \\
&\leq P(\max_{k\leq t\leq k+1} |r_\alpha(B_t^{(\alpha)}) - r_\alpha(B_k^{(\alpha)})| > r_\alpha(\sqrt{n}\varepsilon)) \\
&= P(\max_{k\leq t\leq k+1} |r_\alpha(B_t^{(\alpha)}) - r_\alpha(B_k^{(\alpha)})| > r_\alpha(\sqrt{n}\varepsilon)) \\
&= P(\max_{k\leq t\leq k+1} |r_\alpha(B_t^{(\alpha)}) - r_\alpha(B_k^{(\alpha)})| > (1-\alpha)\sqrt{n}\varepsilon).
\end{aligned}$$

Thus, by Walsh's martingale lemma and Doob's maximal inequalty, one has

$$P(\max_{1\leq k\leq n} \max_{k\leq t\leq k+1} |B_t^{(\alpha)} - B_k^{(\alpha)}| > \sqrt{n}\varepsilon) \leq \sum_{k=1}^{n} \frac{\mathbb{E}|r_\alpha(B_{k+1}^{(\alpha)}) - r_\alpha(B_k^{(\alpha)})|^4}{(1-\alpha)^4 n^2 \varepsilon^4} \leq c(\alpha, \varepsilon)\frac{1}{n}, \tag{25.18}$$

for a positive constant $c(\alpha, \varepsilon)$. It follows that $P(I_n^{(3)} > \varepsilon) \leq c(\alpha, \varepsilon)\frac{1}{n}$. In particular,

$$I_n^{(3)} \leq n^{-\frac{1}{2}} \max_{0\leq k\leq n-1} \max\{|B_t^{(\alpha)} - B_k^{(\alpha)}| : k \leq t \leq k+1\} \to 0 \quad \text{in probability.} \tag{25.19}$$

Hence we need to prove, as $n \to \infty$,

$$I_n^{(1)} := n^{-\frac{1}{2}} \max_{1\leq k\leq n} |B_{T_k}^{(\alpha)} - B_k^{(\alpha)}| \longrightarrow 0 \quad \text{in probability.} \tag{25.20}$$

This follows exactly as for the case of the convergence of simple symmetric random walk to Brownian motion. Namely, since $T_n/n \to 1$ a.s., by the strong law of large numbers, it follows that

$$\varepsilon_n := \max_{1\le k\le n} |\frac{T_k}{n} - \frac{k}{n}| \longrightarrow 0 \quad \text{as } n \to \infty \text{ (almost surely).} \tag{25.21}$$

So, in view of (25.21), there exists for each $\varepsilon > 0$ an integer n_ε such that $P(\varepsilon_n < \varepsilon) > 1-\varepsilon$ for all $n \ge n_\varepsilon$. Hence with probability greater than $1-\varepsilon$ one has for all $n \ge n_\varepsilon$, the estimate

$$\begin{aligned} I_n^{(1)} &\le \max_{\substack{|s-t|\le n\varepsilon,\\ 0\le s,t\le n+n\varepsilon}} n^{-\frac{1}{2}}|B_s^{(\alpha)} - B_t^{(\alpha)}| = \max_{\substack{|s-t|\le n\varepsilon,\\ 0\le s,t\le n(1+\varepsilon)}} |\widetilde{B}_{s/n}^{(\alpha,n)} - \widetilde{B}_{t/n}^{(\alpha,n)}| \\ &= \max_{\substack{|s'-t'|\le \varepsilon,\\ 0\le s',t'\le 1+\varepsilon}} |\widetilde{B}_{s'}^{(\alpha,n)} - \widetilde{B}_{t'}^{(\alpha,n)}| \overset{d}{=} \max_{\substack{|s'-t'|\le \varepsilon,\\ 0\le s',t'\le 1+\varepsilon}} |B_{s'}^{(\alpha)} - B_{t'}^{(\alpha)}| \\ &\longrightarrow 0 \quad \text{as } \varepsilon \downarrow 0, \end{aligned}$$

by the continuity of $t \to B_t^{(\alpha)}$. Given $\delta > 0$ one may then choose $\varepsilon = \varepsilon_\delta$ such that for all $n \ge n(\delta) := n_{\varepsilon_\delta}$, $P(I_n^{(1)} > \delta) < \delta$. Hence $I_n^{(1)} \to 0$ in probability. ■

A practical value of this extension is in the design of numerical schemes[8] for more general models. Hoteit et al. (2002) describes four different probability schemes from the existing literature. The skew random walk convergence provides a theoretical justification for one of these schemes.

Another interesting application that arises in this general context of solute transport involves experiments[9] to determine the so-called breakthrough curves, or passage time distributions, in the presence of an interface. In particular, a distinct skewness was empirically observed in these experiments that was eventually explained as a consequence of skew diffusion.[10]

This application motivated the following explicit computation of the first passage time for skew Brownian motion obtained by Appuhamillage and Sheldon (2012). It involves first characterizing the distribution of ranked excursion heights of skew Brownian motion. The ranked excursion heights are then used to obtain a formula[11] for the first passage time distribution.

Let

$$M_1^{(\alpha)}(t) \ge M_2^{(\alpha)}(t) \ge \ldots \ge 0$$

be the ranked decreasing sequence of excursion heights $\sup_{s\in J_m\cap[0,t]} B_s^{(\alpha)+}$ ranging over all m such that $J_m \cap [0,t]$ is nonempty. Note that $B_s^{(\alpha)+}$ denotes the positive

[8] See Lejay and Martinez (2006), Bokil et al. (2020), and references therein, for more extensive treatments of the numerical problems.

[9] Berkowitz et al. (2009).

[10] See Appuhamillage et al. (2010, 2011).

[11] Appuhamillage and Sheldon (2012) provide refined expressions for these formulae as well.

part of $B_s^{(\alpha)}$, so that a negative excursion has height zero and the height of the final excursion is included in the ranked list even if that excursion is incomplete.

For the purposes of coupling the various skew Brownian motions represented by (25.9), the dependence on $\alpha \in (0, 1)$ of the excursion flips $\{A_m \equiv A_m^{(\alpha)} : m = 0, 1, \ldots\}$ will be made explicit, i.e.,

$$B_t^{(\alpha)} = \sum_{m=1}^{\infty} \mathbf{1}_{J_m}(t) A_m^{(\alpha)} |B_t|. \tag{25.22}$$

A coupling construction: Let $\{A_m^{(\alpha/\beta)}: m = 0, 1, \ldots\}$ be a sequence of i.i.d. ± 1 Bernoulli random variables independent of $A^{(\beta)}$ and B with $P(A_m^{(\alpha/\beta)} = 1) = \alpha/\beta$. Define $A_m^{(\alpha)}$ as follows:

$$A_m^{(\alpha)} = \begin{cases} 1 & A_m^{(\beta)} = 1,\ A_m^{(\alpha/\beta)} = 1, \\ -1 & \text{otherwise.} \end{cases}$$

By construction, the sequence $\{A_m^{(\alpha)} : m = 0, 1, \ldots\}$ consists of i.i.d. ± 1 Bernoulli random variables that are independent of B with $P(A_m^{(\alpha)} = 1) = \alpha$. Hence, by using $A_m^{(\alpha)}, m \geq 1$, as the excursion signs in (25.22), we obtain an α-skew Brownian motion $B^{(\alpha)}$. This is a two-stage construction: first, define $B^{(\beta)}$ by independently setting each excursion of $|B|$ to be positive with probability β; then, for each positive excursion of $B^{(\beta)}$, independently decide whether to keep it positive (with probability α/β), or flip it to be negative (with probability $1 - \alpha/\beta$).

Theorem 25.5 (Appuhamillage and Sheldon (2012)) *Fix $y \geq 0$, $t > 0$ and $\alpha, \beta \in (0, 1)$. For each $j = 1, 2, \ldots$, the following relation between ranked excursion heights of α- and β-skew Brownian motions holds:*

$$P_0(M_j^{(\alpha)}(t) > y) = \sum_{k=1}^{\infty} \binom{k-1}{j-1} (1 - \tfrac{\alpha}{\beta})^{k-j} (\tfrac{\alpha}{\beta})^j P_0(M_k^{(\beta)}(t) > y). \tag{25.23}$$

The proof of Theorem 25.5 involves the following inversion formula (Exercise 10).

Lemma 7 (Pitman and Yor (2001), Lemma 9) *Let*

$$b_k = \sum_{m=0}^{\infty} \binom{m}{k} a_m, \quad k = 0, 1, \ldots$$

be the binomial moments of a nonnegative sequence $(a_m,\ m = 0, 1, \ldots)$. Let $B(\theta) := \sum_{k=0}^{\infty} b_k \theta^k$ and suppose $B(\theta_1) < \infty$ for some $\theta_1 > 1$. Then

$$a_m = \sum_{k=0}^{\infty} (-1)^{k-m} \binom{k}{m} b_k, \quad m = 0, 1, \ldots,$$

where the series is absolutely convergent.

Proof of Theorem 25.5 For $\alpha < \beta$, by the two-stage coupling construction of the excursion sign $A_m^{(\alpha)}$ that $M_j^{(\alpha)}(t) = M_{H_j}^{(\beta)}(t)$, where H_j has a negative binomial distribution:

$$P(H_j = h) = \binom{h-1}{j-1} (1 - \tfrac{\alpha}{\beta})^{h-j} (\tfrac{\alpha}{\beta})^j.$$

Hence

$$P_0(M_j^{(\alpha)}(t) > y) = \sum_{k=1}^{\infty} \binom{k-1}{j-1} (1 - \tfrac{\alpha}{\beta})^{\ell-j} (\tfrac{\alpha}{\beta})^j P_0(M_k^{(\beta)}(t) > y). \tag{25.24}$$

For $\beta < \alpha$, the relation can be inverted by an application of Lemma 7. Let $k := j-1$ and $m := h - 1$. Then one can write (25.23) as

$$P_0(M_{k+1}^{(\alpha)}(t) > y) = \sum_{m=0}^{\infty} \binom{m}{k} (1 - \tfrac{\alpha}{\beta})^{m-k} (\tfrac{\alpha}{\beta})^{k+1} P_0(M_{m+1}^{(\beta)}(t) > y). \tag{25.25}$$

We then apply Lemma 7 to the sequences

$$b_k := (1 - \tfrac{\alpha}{\beta})^k (\tfrac{\alpha}{\beta})^{-k-1} P_0(M_{k+1}^{(\alpha)}(t) > y), \ a_m := (1 - \tfrac{\alpha}{\beta})^m P_0(M_{m+1}^{(\beta)}(t) > y).$$

After simplifying, we obtain

$$P_0(M_j^{(\beta)}(t) > y) = \sum_{h=1}^{\infty} \binom{h-1}{j-1} (1 - \tfrac{\beta}{\alpha})^{h-j} (\tfrac{\beta}{\alpha})^j P_0(M_h^{(\alpha)}(t) > y). \tag{25.26}$$

■

Let

$$T_y^{(\alpha)} = \inf\{s \geq 0 : B_s^{(\alpha)} = y\}$$

denote the first time for α-skew Brownian motion to reach y.

In the case of Brownian motion the following formula was obtained by Csaki and Hu (2003) as a limit of the corresponding results for simple symmetric random walk using reflection principle calculations left as an Exercise 11.

Lemma 8 (Csaki and Hu (2003)) *In the case of standard Brownian motion ($\alpha = 1/2$), one has*

$$P_0(M_j^{(1/2)}(t) > y) = 2(1 - \Phi((2j-1)\frac{y}{\sqrt{t}})), \tag{25.27}$$

where Φ is the standard normal distribution function.

Corollary 25.6 *Fix $y \geq 0$ and $t > 0$. Then for each $j = 1, 2, \ldots$, the distribution of $M_j^{(\alpha)}(t)$ is given by the formula*

$$P_0(M_j^{(\alpha)}(t) > y) = \sum_{h=1}^{\infty} 2\binom{h-1}{j-1}\big(1-2\alpha\big)^{h-j}\big(2\alpha\big)^j\big(1 - \Phi((2h-1)y/\sqrt{t})\big),$$

where $\Phi(\cdot)$ is the standard normal distribution function.

Proof The result is immediate from Theorem 25.5 using Lemma 8 and taking $\beta = 1/2$ in (25.23). ■

Corollary 25.7 *Fix $t > 0$. Then*

$$\begin{aligned} &P_0(T_y^{(\alpha)} \in dt) \\ &= \begin{cases} 2\alpha \sum_{h=1}^{\infty}(1-2\alpha)^{h-1} \frac{(2h-1)y}{\sqrt{2\pi}\ t^{3/2}} \exp\{-\ \frac{((2h-1)y)^2}{2t}\}\, dt,\ y > 0 \\ 2(1-\alpha)\sum_{h=1}^{\infty}(2\alpha-1)^{h-1} \frac{(2h-1)(-y)}{\sqrt{2\pi}\ t^{3/2}} \exp\{-\ \frac{((2h-1)y)^2}{2t}\}\, dt,\ y < 0. \end{cases} \end{aligned}$$

Proof For the case $y > 0$ and $t > 0$, we have the following relation between the distributions of $T_y^{(\alpha)}$ and the highest excursion of skew Brownian motion started at 0:

$$P_0(T_y^{(\alpha)} < t) = P_0(M_1^{(\alpha)}(t) > y).$$

Thus using Corollary 25.6, one has

$$\begin{aligned} P_0(T_y^{(\alpha)} < t) &= P_0(M_1^{(\alpha)}(t) > y) \\ &= 4\alpha \sum_{h=1}^{\infty}(1-2\alpha)^{h-1}\int_{\frac{(2h-1)y}{\sqrt{t}}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{z^2}{2}\}\, dz. \end{aligned}$$

The result is immediate after taking the derivative of the above expression with respect to t. For the case $y < 0$, use the relation $P_0(T_y^{(\alpha)} \in dt) = P_0(T_{-y}^{(1-\alpha)} \in dt)$ and the case $y > 0$. ■

Exercises

1. Compute the first passage time density (breakthrough curves) for one-dimensional Brownian motion, with and without drift.
2. Show that if $\{X_n : n = 0, 1, \dots\}$ is a sequence of random maps with values in a Banach space S such that X_n converges in distribution to X, then for any sequence of random maps $\{Y_n : n = 0, 1, \dots\}$ such that $||Y_n|| \to 0$ in probability as $n \to \infty$, where 0 is the zero vector in S, one has that $X_n + Y_n$ converges in distribution to X. [*Hint*: By Alexandrov's theorem,[12] it suffices to check that $\mathbb{E}f(X_n + Y_n) \to \mathbb{E}f(X)$ as $n \to \infty$ for bounded, uniformly continuous functions $f : S \to \mathbb{R}$.]
3. Verify the formulae for the transition probabilities for skew Brownian motion. [*Hint*: Begin with the calculation of $P_0(B_t^{(\alpha)} > y)$ in the case $x = 0$ and $y > 0$. Next consider the case $x > 0$, $y > 0$ by conditioning on the hitting time $\tau_0 = \inf\{t : B_t = 0\}$ of 0 starting from x. The case $x < 0, y > 0$ is also obtained from the reflection principle. Finally, use symmetry to obtain the case $x > 0, y < 0$ from $x < 0, y > 0$.]
4. Let $\{Y_n : n \geq 0\}$ denote the skew random walk with parameter α starting at zero, and let r_α be the reversal function $r_\alpha(x) = (1-\alpha)x\mathbf{1}_{[0,\infty)}(x) + \alpha x\mathbf{1}_{(-\infty,0]}$. Show that $M_n = r_\alpha(Y_n), n = 0, 1, 2, \dots$ is a martingale.
5. Show that the embedding of simple symmetric random walk in Brownian motion is recursively by waiting for the Brownian motion to escape intervals of unit length centered at respective hitting positions; i.e., starting at $S_0 = B_0 = 0$, $S_1 = \pm 1$ according to $B_{T_1} = \pm 1$, $S_2 = S_1 \pm 1$ according to $B_{T_2} = B_{T_1} \pm 1$, etc.
6. Verify that in the case of constant diffusion coefficient $D^+ = D^- = D$, the solution to (25.1) for twice-continuously differentiable initial data c_0, is given by $c(t, y) = \mathbb{E}_y c_0(Y_t), t \geq 0, y \in \mathbb{R}$, where Y is a Brownian motion with diffusion coefficient D and zero drift.
7. For fixed y, the transition probabilities $p^{(\alpha)}(t; x, y)$ satisfy both the interface condition and continuity in the backward variable x.
8. Verify that for fixed x the transition probabilities $p(t; x, y)$ given by (25.11) satisfy the continuity of flux condition in the forward variable y, and is continuous in the forward variable y. In fact, $p(t; x, y)$ is jointly continuous in the backward and forward variables.
9. Verify Corollary 25.3.
10. (*Pitman and Yor Inversion Lemma*) Prove Lemma 7. [*Hint*: Define $A(z) = \sum_{m=1}^\infty a_m z^m$. Note that $k!b_k = \frac{d^k}{dz^k}A(1)$ so that $A(1+\theta) = B(\theta)$, $|\theta| < \theta_1 - 1$. Now compute $A(z) = B(z-1) = \sum_{k=1}^\infty b_k(z-1)^k$ by the binomial (Taylor) expansion of $(z-1)^k$. Note absolute convergence from $A(2+z) = B(1+z)$.]

[12] See BCPT, p. 137.

11. Let $\{S_n : n \geq 0\}$, $S_=0$ be the simple symmetric random walk. Let $\rho_0 = 0$, $\rho_j = \inf\{k > \rho_{j-1} : S_k = 0\}$, $j \geq 1$ denote the successive return times to zero, $\tau_j = \rho_j - \rho_{j-1}$, $j \geq 1$ the length of time between returns, and $N_n = |\{0 < k \leq n : S_k = 0\}|$ the number of returns to zero by time n. Also let $\mu_j = \max_{\rho_{j-1}<k<\rho_j-1} |S_k|$ and $\mu_j^+ \max_{\rho_{j-1}<k<\rho_j-1} S_k$, for $j \geq 1$ denote heights of the excursion. An excursion $(S_{\rho_{j-1}}, \ldots, S_{\rho_j-1})$ is assigned the sign of the random walk with the excursion. If the jth excursion is negative, then $\mu_j^+ = 0$. To order the excursions by height, the j-th largest term of the sequence $\mu_1, \ldots, \mu_{N_n}, \max_{\rho_{N_n} \leq k \leq n} |S_k|$ is denoted $M^{(j)}(n)$, and the j-th largest term of the sequence $\mu_1, \ldots, \mu_{N_n}, \max_{\rho_{N_n} \leq k \leq n} S_k$ is denoted $M_+^{(j)}(n)$. Also define $M_{(n)}^{(j)} = M_+^{(j)}(n) = 0$ if $j > N_n + 1$. Let $y > 0$, ℓ be integers. Use successive applications of the reflection principle[13] to show that

(i) $P(M_+^{(j)}(n) \geq y, S_n = \ell) = P(S_n = 2jy - \ell), \quad \ell < y.$
(ii) $P(M_+^{(j)}(n) \geq y, S_n = \ell) = P(S_n = 2(j-1)y + \ell), \quad \ell \geq y.$
(iii) $P(M_+^{(j)}(n) \geq y) = P(S_n = (2j-1)y) + 2P(S_n > (2j-1)y).$

[*Hint*: The event $[M_+^{(j)}(n) \geq y]$ implies at least two upcrossings from 0 to y by time n. Consider parts of the process between first hitting time of y, time of first return to 0, the next hitting time of y after returning to 0, and cases of $\ell < y, \ell \geq y$, respectively, to construct the indicated bijections. Apply the central limit theorem and Stirling's formula to obtain the formula given in Lemma 8.]

[13] This is an approach to Lemma 8 used by Csaki and Hu (2003). This is not unlike the use of the method of images in Bhattacharya and Waymire (2021), Cor. 3.4, p. 30.

Chapter 26
Special Topic: Piecewise Deterministic Markov Processes in Population Biology

In this chapter a stochastic process is introduced as a model of deterministic population growth subject to independent random disturbances that either reduce or increase the population size at the arrival times of a homogeneous Poisson process. This is a special case of the general class of piecewise deterministic Markov processes. The central questions concern persistence and evolution toward a steady state.

There is a long history[1] of deterministic sigmoidal growth curves $G(N)$ for population models

$$\frac{dN}{dt} = G(N(t)). \tag{26.1}$$

Models with a linear lowest-order term typically exhibit exponential growth at early times. The Richards growth model[2] $G = G_\theta$, where

$$G_\theta(N) = rN\left[1 - (N/k)^\theta\right], \quad \theta > 0, \tag{26.2}$$

is a general model with this feature, generalizing the logistic and exponential models as special cases, and will be assumed throughout the treatment presented

[1] Kingsland (1982) provides a very nice summary of this history with extensive references. Also see Bacaer (2011) for the special case of logistic growth.

[2] The Richards (1959) model is also known as the *theta-logistic* model Lande et al. (2003); Gilpin and Ayala (1973).

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8_26

in this chapter. Here the parameter k is referred to as the *carrying capacity* of the population. Note from (26.1) that the population size N increases for $N < k$ and decreases for $N > k$, with a unique maximum at $N = k$.

While these and other classical models differ in details, the solutions share a common *averaging dynamic* that is noteworthy. Namely, a suitably transformed[3] measure of the population size evolves as a temporally weighted average between the (transformed) initial population size and the (transformed) carrying capacity. The proof rests on a more basic fact that every increasing, continuously differentiable function $x(t)$, $0 \le t < \infty$, can be represented as a weighted average by time-varying weights of its initial data x_0 and its asymptotic limit $x_\infty < \infty$ (assumed finite for simplicity) as follows:

Lemma 1 *Suppose that* $x : [0, \infty) \to [x_0, x_\infty)$ *is continuously differentiable and monotonically increasing with* $x_\infty = \lim_{t\to\infty} x(t) < \infty$*, and let* $\nu > 0$ *be arbitrary. Then there exists a monotone (increasing or decreasing) function* $h : [x_0, x_\infty] \to \mathbb{R}$ *such that*

$$h(x(t)) = h(x_\infty)(1 - e^{-\nu t}) + h(x_0)e^{-\nu t}, \ t \ge 0.$$

Proof Since $x(t)$ is continuously differentiable and increasing, it is invertible with continuously differentiable inverse $\tau(x)$, where $\tau : [x_0, x_\infty) \to [0, +\infty)$, such that $\tau(x_0) = 0$, $\lim_{x\to x_\infty} \tau(x) = +\infty$, and $d\tau/dx > 0$. Let c_1 and c_2 be arbitrary real numbers such that $c_1 \ne c_2$. Define $h : [x_0, x_\infty] \to \mathbb{R}$ as follows:

$$h(x) := c_1(1 - e^{-\nu\tau(x)}) + c_2 e^{-\nu\tau(x)} \tag{26.3}$$

Then h is continuously differentiable, with derivative

$$\frac{dh}{dx} = (c_1 - c_2)\nu \frac{d\tau}{dx}.$$

Since $c_1 \ne c_2$, $\nu > 0$ and $d\tau/dx > 0$, it follows that h is increasing (if $c_1 > c_2$) or decreasing (if $c_1 < c_2$). In either case, h is invertible, and by taking the inverse in the definition (26.3) and setting $x = x(t)$, we obtain the claimed result by noting that $c_1 = h(x_\infty)$ and $c_2 = h(x_0)$ (recall that $\tau(x_0) = 0$ and $\lim_{x\to x_\infty} \tau(x) = +\infty$). ∎

Such averaging dynamics explicitly reveal the (non-transformed) function in Lemma 1 to be represented by

$$x(t) = h^{-1}\left(h(x_\infty)(1 - e^{-\nu t}) + h(x_0)e^{-\nu t}\right), \ t \ge 0. \tag{26.4}$$

[3] One-to-one transformations, e.g., logarithms, exponentials, reciprocals, centering and scalings, are often found to be convenient when graphing and analyzing biological data.

In particular, if h is such that for all $N(0) = N_0 \in (0, k]$,

$$h(N(t)) = h(k)(1 - e^{-\nu t}) + h(N_0)e^{-\nu t}, \ t \geq 0,$$

then $x(t) = h(N(t))$ will solve the affine-linear equation

$$\begin{aligned}\frac{dx(t)}{dt} &= \frac{d}{dt}\{h(k)(1 - e^{-\nu t}) + h(N_0)e^{-\nu t}\} \\ &= -\nu\{h(k)(1 - e^{-\nu t}) + h(N_0)e^{-\nu t} - h(k)\} \\ &= -\nu h(x(t)) + \nu h(k), \end{aligned} \tag{26.5}$$

for all $t \geq 0$ and all $N_0 \in (0, k]$. The converse holds as well.

This simple realization will be useful for analyzing the disturbance model to be introduced below (see (26.8)). In particular, for Richards growth (also see Exercise 5),

$$h(N) = 1/N^\theta$$

transforms the nonlinear equation (26.1) (for (26.2)) into the affine linear equation (26.5) with $\nu = \theta r$. The solution of (26.1), (26.2) is given by

$$\frac{1}{N^\theta(t)} = \frac{1}{k^\theta}(1 - e^{-\theta r t}) + \frac{1}{N_0^\theta}e^{-\theta r t}, \tag{26.6}$$

or equivalently, $N^\theta(t)$ evolves as a *harmonic average* of k^θ and N_0^θ of the form

$$N^\theta(t) = \frac{1}{\frac{1}{k^\theta}(1 - e^{-\theta r t}) + \frac{1}{N_0^\theta}e^{-\theta r t}}. \tag{26.7}$$

Next consider the possibility of episodic disturbances of the population, i.e., *random catastrophes*. Familiar examples of catastrophes are numerous, including severe storms, meteor impacts, epidemics, forest fires, floods, droughts, infestations, volcanic eruptions, and so on. The episodic nature of such disturbances means that it is more natural to represent their occurrence times as a Poisson process in time with intensity parameter λ determining the mean frequency of occurrence. One can model the resulting mortality either by subtracting a random number from the population or by assuming that only a random fraction of the population survives the disturbance. However, the latter multiplicative model is natural in this case since it scales with the population size, i.e., the mortality in an additive model, can be larger than the total size of the population. Note that whether the mortality due to a

catastrophe is additive[4] or multiplicative, the resulting stochastic process for the population size, $N(t)$, is no longer continuous but rather piecewise continuous, with jump discontinuities occurring at the times of catastrophes.[5]

When treating disturbance models, the blanket assumption that the disturbance factors are positive with positive probability is made throughout. Obviously if disturbances $\mathcal{D}$ are permitted to destroy the entire population with probability one, i.e., $P(\mathcal{D} = 0) = 1$, then there can be no recovery for the given model.

The stochastic process of interest here falls within a general class of piecewise deterministic Markov processes[6] singled out by Davis (1984),[7] in which a single-species population undergoes deterministic growth determined by an ordinary differential equation (26.1) but which also experiences random, episodic disturbances that remove a random fraction of the population. In the present model, net growth is deterministic, while the frequency and magnitude of disturbances that lead to mortality are treated as stochastic. The competition between the population's net reproductive rate and its mortality rate due to disturbances sets up a situation where critical thresholds can be computed in terms of model parameters that determine what will happen to the size of the population in the long term. This model can be expressed more precisely as follows:

$$\begin{aligned}
&\frac{dN}{dt}(t) = G(N(t)), \quad \tau_i \le t < \tau_{i+1}, \quad i = 0, 1, 2, \ldots, \\
&N(\tau_i) = \mathcal{D}_i\, N(\tau_i^-), \quad N(0) = N_0 > 0, \text{ with } N_0 < k,
\end{aligned} \tag{26.8}$$

where $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$ is the sequence of arrival times of a Poisson renewal process $\{\Lambda(t) : t \ge 0\}$ with intensity $\lambda > 0$ and $\mathcal{D}_1, \mathcal{D}_2, \ldots$ is a sequence of independent and identically distributed (i.i.d.) disturbance factors on the interval [0, 1] and independent of the arrival time process. These disturbance factors determine the fraction of the population that survives a given disturbance. The Richards growth curve $G = G_\theta$ will be considered for specificity of results.[8]

One may note that $G_\theta(0) = 0$ implies that $N = 0$ is an absorbing state for (26.8). In particular the Dirac (point mass) probability distribution δ_0 is an

[4] See Chapter 11, Example 1 for results in the case of additive perturbations of the population. Also see Goncalves et al. (2022), Schlomann (2018) for alternative models.

[5] Hanson and Tuckwell (1978, 1981, 1997) appear to be among the earliest to consider population dynamics models that included random catastrophes. In each of their papers, these authors modeled the disturbance times with a Poisson event process and represented the growth between disturbances by deterministic, logistic growth.

[6] This class of models gained attention in Bayesian statistics, e.g., Azais and Bouguet (2018), cellular biology, e.g., Crudu et al. (2012), micro-insurance, e.g., Henshaw (2022), and fluid flows, e.g., Majda and Tong (2016), to name a few other applications outside of the present chapter.

[7] Davis (1984) provides a complete characterization of the piecewise deterministic Markov processes in terms of the general form of their infinitesimal generator.

[8] More general considerations involving the qualitative features of the growth curves are included in Peckham et al. (2018).

invariant (equilibrium) distribution. In particular, (26.8) defines a *reducible* Markov process in the sense that the state $N = 0$ is inaccessible from states in $(0, \infty)$. We are interested in conditions under which this is the only invariant distribution, as well as conditions in which another invariant distribution also exists on the interval $(0, k)$. Of course, having two distinct invariant distributions on $[0, k)$ means that there will be infinitely many since randomizing between the two is also invariant.

In addition to the continuous time model (26.8), a natural discrete time model is obtained by considering the population sizes at the sequence of times at which disturbances occur. That is,

$$N_n = \mathcal{D}_n N(\tau_n^-), \quad n = 0, 1, 2, \dots, \quad \tau_0^- = 0, \quad X_0 = 1, \tag{26.9}$$

where N_n is the random size of the population immediately *after* the nth episodic disturbance. The left-hand limit

$$N(\tau_n -) = \lim_{t \uparrow \tau_n} N(t)$$

refers to the population size just *before* the nth disturbance. That such discrete parameter Markov processes may be viewed as i.i.d. iterated maps[9] can be used to great advantage, especially for non-irreducible Markov processes, will naturally emerge for this application.

Theorem 26.1 (Threshold for Convergence in Distribution) *Let τ_n be a sequence of arrival times of a Poisson process with intensity $\lambda > 0$ and $\mathcal{D}_n$ be a sequence of i.i.d. random disturbance variables on* $[0, 1]$ *which is independent of the Poisson process. Consider $G_\theta(N) = rN(1 - (N/k)^\theta)$ for some $r > 0$, $\theta > 0$, and $k > 0$.*

a. *If $\mathbb{E}[\ln(\mathcal{D}_1)] + \frac{r}{\lambda} > 0$, then $\{N_n\}_{n=0}^{\infty}$ converges in distribution to a unique invariant distribution with support on* $(0, k)$.

b. *If $\mathbb{E}[\ln(\mathcal{D}_1)] + \frac{r}{\lambda} < 0$, then $\{N_n\}_{n=0}^{\infty}$ converges in distribution to* 0. *Moreover, in this case, the convergence to $\delta_{\{0\}}$ is exponentially fast in the (Prokhorov) metric of convergence in distribution.*

Proof The proof relies on some ergodic theory for discrete parameter Markov processes represented as i.i.d. iterated random maps in rather interesting ways. The specific conditions employed to prove the existence of unique invariant probabilities can be found in from Bhattacharya and Waymire (2022), Theorem 18.6, Bhattacharya and Majumdar (2007), Theorems 3.1 and 7.1, or in Diaconis and Freedman (1999). First, let us note that the discrete-time disturbance model associated with (26.7) is

[9] See Bhattacharya and Waymire (2022), Bhattacharya and Majumdar (2007), Diaconis and Freedman (1999), Schreiber (2012) for a more general perspective.

$$N_n = \mathcal{D}_n \frac{1}{(\frac{1}{k^\theta}(1-e^{-r\theta T_n}) + \frac{1}{N_{n-1}^\theta} e^{-r\theta T_n})^{1/\theta}}, \qquad (n \geq 1)\,. \tag{26.10}$$

The reciprocal transform for the analysis of the Richards growth model is given by $J_n = 1/N_n^\theta \in (k^{1-\theta}, \infty)$. Define $Y_n = \exp(-r\theta\, T_n)$, $A_n = Y_n \mathcal{D}_n^{-\theta}$, and $B_n = (1 - Y_n)\, k^{-\theta} \mathcal{D}_n^{-\theta}$. One then has

$$J_n = A_n\, J_{n-1} + B_n, \qquad J_0 = 1/N_0^\theta. \tag{26.11}$$

From (26.11) it follows from the ergodic theory (cited above) for iterated affine random maps on the complete and separable metric space $[\frac{1}{k}, \infty)$ that under the condition that

$$\mathbb{E} \ln A_1 < 0, \quad \mathbb{E}(\ln B_1)^+ < \infty, \tag{26.12}$$

$J_n = 1/N_n^\theta$, evolves to a unique, invariant distribution on $[k^{-\theta}, \infty)$. These conditions are readily checked. From this the assertion (a) follows because the map $x \to 1/x^\theta$ of $(0, k]$ onto $[k^{-\theta}, \infty)$ is continuous with a continuous inverse.

The recursion (26.10) can also be directly expressed as iterated random function dynamics,

$$N_n = \gamma_{\mathcal{A}^{(n)}} \circ \gamma_{\mathcal{A}^{(n-1)}} \circ \cdots \circ \gamma_{\mathcal{A}^{(1)}}(N_0), \tag{26.13}$$

where the random vectors

$$\mathcal{A}^{(i)} = (\alpha_1^{(i)}, \alpha_2^{(i)}), \quad i \geq 1,$$

are i.i.d. with independent components such that $\alpha_1 = \mathcal{D}_1 \in (0, 1)$, α_2 is exponentially distributed with parameter $\lambda > 0$, and

$$\gamma_{\mathcal{A}}(x) \equiv \gamma_{(\alpha_1,\alpha_2)}(x) = \alpha_1 \frac{1}{\frac{1}{k^\theta}(1 - e^{-r\alpha_2}) + \frac{1}{x^\theta} e^{-r\alpha_2}}. \tag{26.14}$$

To prove (b) observe that these iterated maps are Lipschitz maps on the complete and separable metric space $[0, k]$ with Lipschitz coefficient $\alpha_1 e^{r\alpha_2}$. Moreover $\delta_{\{0\}}$ is the unique invariant probability, with exponential convergence, under the condition (cited above) that

$$0 > \mathbb{E} \ln(\alpha_1 e^{r\alpha_2}) = \mathbb{E} \ln \mathcal{D}_1 + \frac{r}{\lambda}.$$

■

Remark 26.1 While the threshold condition does not depend on the parameter θ, when it exists, the details of the asymptotic distribution will depend on θ.

The following theorem describes the relationship between steady state distributions of the continuous and discrete time evolutions.[10] Let us review that between disturbances, the deterministic law of evolution of the population continuously in time is given by an equation of the general form

$$\frac{dN(t)}{dt} = G_\theta(N(t)), \quad N(0) = x, \tag{26.15}$$

whose solution from (26.6) may be expressed as

$$N(t) = g_\theta(t, x), t \geq 0, x > 0,$$

where the population flows $x \to g_\theta(t, x)$ are continuous, one-to-one maps with a continuous inverse, such that $g_\theta(0, x) = x$, and $g_\theta(s+t, x) = g_\theta(t, g_\theta(s, x)), s, t \geq 0, x > 0$. In particular, the uninterrupted evolutions considered here have unique solutions at all times for a given initial value. Also, the discrete-time post-disturbed population model is given by

$$N_0 = x, \quad N_n = \mathcal{D}_n g_\theta(T_n, N_{n-1}),\ n = 1, 2, \ldots. \tag{26.16}$$

Theorem 26.2 (Continuous and Discrete Time Invariant Distributions) *Let $g_\theta(t, x)$ be the flow of the deterministic system (26.15). Then*

a. *Given an invariant distribution π for the discrete time post-disturbance population model (26.16), let Y be a random variable with distribution π, and let T be an exponentially distributed random variable with parameter λ, independent of Y. Then*

$$\mu(C) = P(g_\theta(T, Y) \in C), \quad C \subset (0, \infty),$$

is an invariant probability measure for the corresponding continuous time disturbance model (26.8).

b. *Given an invariant probability measure μ, for the continuous time disturbance model (26.8), let Y be a random variable with distribution μ, and let X be distributed as the random disturbance factor $\mathcal{D}_1$ in $(0, 1)$, independent of Y. Then*

$$\pi(C) = P(XY \in C), \quad C \subset (0, \infty),$$

[10] This result follows as a special case of a much more general theory for piecewise deterministic Markov processes given by Costa (1990). A proof for a more general class of growth rates is also provided in Peckham et al. (2018).

is an invariant distribution for the corresponding discrete time post-disturbance model (26.16).

Proof For an invariant probability μ, continuous time evolution can be expressed in terms of the semigroup of linear contraction operators defined on $L^2(\mu)$ by

$$T(t)f(x) = \mathbb{E}_x f(N(t)), \quad t \geq 0, x > 0,$$

via its infinitesimal generator given by

$$Af(x) = \frac{d}{dt} f(g_\theta(t,x))|_{t=0} + \lambda\{\mathbb{E}f(Xx) - f(x)\}.$$

To derive this simply observe that up to $o(t)$ error as $t \downarrow 0$, either one or no disturbance will occur in the time interval $[0, t)$. Thus

$$\frac{T(t)f(x) - f(x)}{t} = \frac{f(g_\theta(t,x))e^{-\lambda t} - f(x)}{t} + \frac{1}{t}\int_0^t \mathbb{E}\left(f(Xg_\theta(s,x))\right)\lambda e^{-\lambda s} ds + o(t).$$

The first term is, by the product differentiation rule,

$$\begin{aligned} \frac{f(g_\theta(t,x))e^{-\lambda t} - f(g_\theta(0,x))e^{-\lambda 0}}{t} &\to \frac{d}{dt} f(g_\theta(t,x))e^{-\lambda t}|_{t=0} \\ &= \frac{d}{dt} f(g_\theta(t,x))|_{t=0} - \lambda f(x). \end{aligned}$$

The second term is $\lambda \mathbb{E}f(Xx)$ in the limit as $t \downarrow 0$.

Since μ is an invariant probability for this continuous time evolution, then one has from the adjoint (Fokker-Planck) equation $A^*\mu = \frac{d}{dt}\mu = 0$ for the adjoint operator. In particular, for f belonging to the domain of A as an (unbounded) operator on $L^2(\mu)$,

$$0 =< f, A^*\mu >=< Af, \mu >= \int_0^\infty Af(x)\mu(dx),\ f \in \mathcal{D}_A \subset L^2(\mu).$$

In the case of the discrete time evolution, the one-step transition operator is defined by

$$Mf(x) = \mathbb{E}f(Xg_\theta(T,x)),\ x > 0.$$

The condition for π to be an invariant probability for the discrete time evolution is that for $f \in L^2(\pi) \subset L^1(\pi)$,

$$\int_0^\infty Mf(x)\pi(dx) = \int_0^\infty f(x)\pi(dx).$$

In particular, it suffices to consider indicator functions $f = 1_C, C \subset (0, \infty)$, in which case one has

$$\int_0^\infty P(Xg_\theta(T, x) \in C)\, \pi(dx) = \pi(C).$$

These are the essential calculations required for the proofs of (a) and (b).

Let's begin with part (a). First note from the definition of μ that

$$\int_0^\infty Af(x)\mu(dx) = \int_0^\infty \int_0^\infty Af(g_\theta(t, y))\lambda e^{-\lambda t}\, dt\, \pi(dy).$$

Now, in view of the above calculation of A, one has

$$\int_0^\infty Af(g_\theta(t, y))\lambda e^{-\lambda t}\, dt = \int_0^\infty (\frac{\partial f(g_\theta(t, x))}{\partial t} + \lambda[\mathbb{E} f(Xg_\theta(t, x)) - f(g_\theta(t, x)])\lambda e^{-\lambda t} dt.$$

After an integration by parts, this yields

$$\int_0^\infty Af(g_\theta(t, y))\lambda e^{-\lambda t} dt = \lambda\{\mathbb{E} f(Xg_\theta(T, x)) - f(x)\}$$

Thus, using this and the invariance of π for the discrete process, one has

$$\int_0^\infty Af(x)\mu(dx) = \lambda \int_0^\infty \{\mathbb{E} f(Xg_\theta(T, x)) - f(x)\}\pi(dx) = 0.$$

This proves part (a).

To prove part (b), first apply A to the function $x \to P(Xg_\theta(T, x) \in C)$. First note from the composition property and an indicated change of variable

$$P(Xg_\theta(T, x) \in C) = P(Xg_\theta(T + t, x) \in C) = e^{\lambda t} \int_t^\infty P(Xg(s, x) \in C)\lambda e^{-\lambda s} ds.$$

In particular the first term of $AP(Xg_\theta(T, x) \in C)$ is

$$\frac{d}{dt} P(Xg_\theta(T, x) \in C)|_{t=0} = \lambda\{P(Xg_\theta(T + t, x) \in C) - P(Xx \in C)\}.$$

Adding this to the second term yields

$$AP(Xg_\theta(T, x) \in C) = \lambda\{\int_0^\infty P(Xg_\theta(T, y) \in C)P(Xx \in dy) - P(Xx \in C)\}.$$

Integrating with respect to the continuous time invariant distribution μ yields

$$0 = \lambda \int_0^\infty \{\int_0^\infty P(Xg_\theta(T, y) \in C)P(Xx \in dy) - P(Xx \in C)\}\mu(dx),$$

or equivalently

$$\int_0^\infty \int_0^\infty P(Xg_\theta(T, y) \in C)P(Xx \in dy)\mu(dx) = \int_0^\infty P(Xx \in C)\mu(dx).$$

But since by definition $\pi(dy) = \int_0^\infty P(Xx \in dy)\mu(dx)$, this is precisely the condition

$$\int_0^\infty P(Xg_\theta(T, y) \in C)\pi(dy) = \pi(C),$$

i.e., that π is an invariant probability for the discrete time distribution. ■

The following corollary demonstrates the relationship between invariant distributions of the discrete-time and continuous time stochastic processes as stated in general in Theorem 26.2 for the Richards growth model.

Corollary 26.3 *Assume that the conditions of Theorem 26.1 hold and that*

$$1/p := r/\lambda > -\mathbb{E} \ln \mathcal{D}_1.$$

Then the rescaled continuous time disturbed Richards model $\frac{N}{k}$ *has the invariant cumulative distribution function*

$$\mu_k(0, x] = \int_0^x (\frac{y^{-\theta} - x^{-\theta}}{y^{-\theta} - 1})^{\frac{\lambda}{\theta r}} \pi_k(dy) \quad 0 \le x \le 1, \tag{26.17}$$

where π_k *is the rescaled invariant distribution for the discrete-time distributed Richards model from (a) in Theorem 26.1.*

Proof Assume $r/\lambda > -\mathbb{E}(\ln \mathcal{D}_1)$. Let $\pi_k(dx)$ denote the invariant distribution for the discrete-time post-disturbance Richards model, rescaled by k to a distribution on $[0, 1]$. That is if Y denotes a random variable with distribution $\pi(dx)$ on $[0, k]$, then let $Y_k = k^{-1}Y$ and denote its distribution by $\pi_k(dx)$ on $[0, 1]$. Scaling the post-disturbance evolution accordingly, one has

$$\frac{N_n}{k} = \mathcal{D}_n[e^{-r\theta t} + \left(\frac{N_{n-1}}{k}\right)^{-\theta} (1 - e^{-r\theta T})]^{-\frac{1}{\theta}}.$$

Thus, letting

$$g_\theta(T, Y) = [e^{-r\theta T} + Y_k^{-\theta}(1 - e^{-r\theta T})]^{-\frac{1}{\theta}},$$

where T is exponentially distributed with parameter $\lambda > 0$ and independent of Y_k, by Theorem 26.2 the invariant distribution of the population size rescaled by $\frac{1}{k}$ can be computed as follows:

$$\begin{aligned} P\,(g_\theta(T, Y_k) \le x) &= P(T \ge -\frac{1}{\theta r}\ln(\frac{x^{-\theta} - Y_k^{-\theta}}{1 - Y_k^{-\theta}}), Y_k \le x) \\ &= \mathbb{E}[\mathbf{1}_{[Y_k \le x]}(\frac{x^{-\theta} - Y_k^{-\theta}}{1 - Y_k^{-\theta}})^{\frac{\lambda}{\theta r}}] \\ &= \int_0^x (\frac{y^{-\theta} - x^{-\theta}}{y^{-\theta} - 1})^{\frac{\lambda}{\theta r}} \pi_k(dy) \quad 0 \le x \le 1. \end{aligned}$$

■

Corollary 26.4 *Assume a uniformly distributed disturbance on* $[0, 1]$ *and* $\theta = 1$. *Then the invariant distribution function for the (rescaled) population in the continuous time Richards (logistic) growth model is given by the Beta distribution*

$$\mu_k[0, x] = C_p x^{1-p}, \quad 0 \le x \le q,\ p = \frac{\lambda}{r}.$$

Proof For uniformly distributed disturbances on $[0, 1]$ and $\theta = 1$, one has

$$1/p = r/\lambda > 1 = -\mathbb{E}\ln \mathcal{D}_1.$$

Thus the discrete parameter invariant probability exists, and one may compute as in the previous corollary that Y_k has the pdf $C'_p\,(1 - y)^p\, y^{-p}, 0 \le y \le 1$. From here one has

$$\begin{aligned} \mu_k[0, x] &= \int_0^x (\frac{y^{-1} - x^{-1}}{y^{-1} - 1})^p\, C'_p(1 - y)^p y^{-p}\, dy \\ &= C_p\, x^{1-p}, \quad 0 \le x \le 1,\ p = \frac{\lambda}{r}. \end{aligned}$$

■

A natural extension of the disturbances introduced in this chapter would allow for climatic effects that could produce a gradual increase in the average frequency of disturbances in the Poisson process. That is, $\Lambda(t), t \ge 0$, would be replaced by a time-inhomogeneous Poisson process with a nondecreasing intensity function $\lambda(t), t \ge 0$, e.g., $\lambda(t) = t^\theta$, or $\lambda(t) = \log(1 + t), t \ge 0$. However, if

$\int_0^\infty \lambda(t)dt = \infty$, then it is possible to homogenize the Poisson process by an appropriate (nonlinear) change in time scale (Exercise 3). Under such a time change, the general form of the model remains the same, but the parameters of the logistic curve then become time-dependent. Another way in which the model may be modified is to permit immigrations into the population after the catastrophic events. Such modifications could make interesting projects for further analysis.

Exercises

1. For simple exponential growth, $G_\infty(N) = rN$ for some $r > 0$, $N(\tau_n^-) = N_{n-1}\, e^{r\,T_n}$

   (i) Show that $N_n = N_0 \prod_{k=1}^n [e^{r\,T_k}\mathcal{D}_k]$. where N_0 is a given initial condition in $(0, k)$, $T_n \equiv \tau_n - \tau_{n-1}$ $(n > 1)$ is the random time interval between disturbances and $\mathcal{D}_n$ is the fraction of the population that survives the nth disturbance.

   (ii) Show that $Y_n = \exp(r\,T_n)$ takes values in $[1, \infty)$ and has a Paraeto distribution with cumulative distribution function $F_{Y_n}(s) = 1 - s^{-p}$, $s > 0$, where $p := \lambda/r > 0$. Here, $\mathbb{E}Y_n = p/(p-1) = (\lambda/r)/((\lambda/r) - 1)$ if $\lambda/r > 1$ and is infinite otherwise.

2. (*Disturbed Exponential Growth Model*) Let τ_n be a sequence of arrival times of a Poisson process with intensity $\lambda > 0$, and $\mathcal{D}_n$ be a sequence of i.i.d. random disturbance variables on $[0, 1]$ which is independent of the Poisson process. Consider $G_\infty(N) = rN$ for some $r > 0$.
   (a) Show

   (i) If $0 < \mathbb{E}[\ln \mathcal{D}_1] + \frac{r}{\lambda} < \infty$, then $N_n \to \infty$ a.s. as $n \to \infty$.

   (ii) If $-\infty < \mathbb{E}[\ln \mathcal{D}_1] + \frac{r}{\lambda} < 0$, then $N_n \to 0$ a.s. as $n \to \infty$.

   (b) Show that $\mathbb{E}\,(N_n) \to \begin{cases} 0, & \text{if } \mathbb{E}(\mathcal{D}_1) + \frac{r}{\lambda} < 1 \\ N_0, & \text{if } \mathbb{E}(\mathcal{D}_1) + \frac{r}{\lambda} = 1 \\ \infty, & \text{if } \mathbb{E}(\mathcal{D}_1) + \frac{r}{\lambda} > 1 \end{cases}.$

   [*Hint*: Consider the recursive equations for the expected values.]

3. Assume $\int_0^\infty \lambda(t)dt = \infty$. Show that it is possible to homogenize the Poisson process to a unit rate Poisson process by an appropriate (nonlinear) change in time scale.

4. For the logistic model ($\theta = 1$), show that

$$\mathbb{E}\frac{1}{N_n} \to \begin{cases} \frac{\mathbb{E}\mathcal{D}_1^{-1}}{k(1-\frac{\lambda}{r}(\mathbb{E}\mathcal{D}_1^{-1})-1))}, & \text{if}\, \mathbb{E}\left(\mathcal{D}_1^{-1}\right) - \frac{r}{\lambda} < 1 \\ \infty, & \text{if}\, \mathbb{E}\left(\mathcal{D}_1^{-1}\right) - \frac{r}{\lambda} \geq 1. \end{cases}$$

5. (*Gompertz model*)

   (i) Show that for the Gompertz[11] growth curve, $G_0(N) = -rN\ln(N/k)$, the map $h(N) = \ln(N)$ transforms (26.1) into (26.5) with $\nu = r$.

   (ii) Show that the solution to (26.1) is $\ln N(t) = \ln k(1 - e^{-rt}) + \ln N_0 e^{-rt}$.

   (iii) Show that the Gompertz growth G_0 is obtained in the limit as $\theta \to 0$ from Richards G_θ.

   (iv) Show that the discrete-time disturbance model associated with the Gompertz model can be expressed as $N_n = k(\frac{N_{n-1}}{k})^{e^{-rT_n}}\mathcal{D}_n,\ n \geq 1$.

   (v) Show that $h(N) = \ln(N)$ transforms the (rescaled) Gompertz model as the affine linear model $J_n = A_n J_{n-1} + B_n, \quad J_0 = \ln(N_0/k),\ n = 1, 2, \ldots,$ where $J_n = \ln(N_n/k) \in (-\infty, 0)$ and defining $A_n = e^{-rT_n}$ and $B_n = \ln(\mathcal{D}_n)$.

   (vi) Assume $\mathbb{E}[\ln(-\ln(\mathcal{D}_n))]^+ < \infty$, where $[x]^+ = \max(0, x)$. Show that $\{N_n\}_{n=0}^{\infty}$ converges in distribution to a unique invariant distribution supported on $(0, k)$, i.e., absence of threshold.

   (vii) Assume uniformly distributed disturbances on $[0, 1]$, and let $\tilde{N} = \lim_{n\to\infty}(N_n/k)$ (in distribution). Show that the cdf $F_{\tilde{N}}$ of $\tilde{N}$ is given by $F_{\tilde{N}}(z) = \frac{\Gamma(1+p, -\ln(z))}{\Gamma(1+p)}, \quad z \in (0, 1)$, where the numerator is the incomplete Gamma function. Show that the corresponding pdf is given by $f_{\tilde{N}}(z) = \frac{[-\ln(z)]^p}{\Gamma(1+p)}, \quad z \in (0, 1)$.

   (viii) Assuming uniformly distributed disturbances on $[0, 1]$, show that $\mathbb{E}(\tilde{N}^m) = (1 + m)^{-(1+p)}$, for $m > -1$

[11] This growth curve dates back to Gompertz (1825).

Appendix A
The Hille–Yosida Theorem and Closed Graph Theorem

For simplicity, we assume the Banach space $\mathscr{X}$ in the next theorem to be real. See Definitions 2.3, 2.4, and 2.7 for context.

Theorem A.1 (Hille–Yosida Theorem) *A linear operator A on $\mathscr{X}$ is the infinitesimal generator of a strongly continuous contraction semigroup if and only if (1) A is densely defined and (2) for all $\lambda > 0$, $\lambda \in \rho(A)$, and $||R_\lambda|| \leq 1/\lambda$.*

Proof The necessity proof ("only if") is contained in Theorem 2.7. **S**ufficiency. Consider the resolvent operator R_n for $n = 1, 2, \ldots$. Since $(n - A)R_n g = g = R_n(n - A)g$ $\forall g \in D_A$,

$$AR_n g = R_n A g \qquad \forall g \in D_A \tag{A.1}$$

Also, AR_n is a bounded operator. Indeed, $(n - A)R_n = I$ implies, by hypothesis,

$$AR_n = nR_n - I. \qquad (||AR_n|| \leq ||nR_n|| + ||I|| \leq 2). \tag{A.2}$$

Now $\forall g \in D_A$, $R_n(n - A)g = g$ implies

$$||(nR_n - I)g|| = ||R_n A g|| \leq \frac{1}{n}||Ag|| \to 0.$$

Since $||nR_n - I|| \leq 2$ and D_A is dense in $\mathscr{X}$, the last relation implies

$$||nR_n g - g|| \to 0. \qquad \forall g \in \mathscr{X}. \tag{A.3}$$

Now define (see (A.1))

$$A^{(n)} = nAR_n = n(nR_n - I). \qquad (n = 1, 2, \ldots) \tag{A.4}$$

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8

Then $A^{(n)}$ is a *bounded* operator ($||A^{(n)}|| \leq n||nR_n - I|| \leq 2n$) and, by (A.1) and (A.3),

$$A^{(n)} f = nAR_n f = nR_n Af \to Af \qquad \forall f \in D_A. \tag{A.5}$$

For each $n = 1, 2, \ldots$ consider the semigroup $\{e^{tA^{(n)}} : 0 \leq t < \infty\}$. This is a contraction:

$$\begin{aligned} ||e^{tA^{(n)}}|| &= ||e^{tn(nR_n - I)}|| = ||e^{-nt} \cdot e^{n^2 R_n}|| \\ &\leq e^{-nt} e^{||tn^2 R_n||} \leq e^{-nt} \cdot e^{nt} = 1. \end{aligned}$$

Now one has, using the lemma below:

$$||e^{tA^{(n)}} f - e^{tA^{(m)}} f|| \leq t||A^{(n)} f - A^{(m)} f|| \qquad (\forall f \in \mathscr{X}).$$

Combining this with (A.5), one gets

$$||e^{tA^{(n)}} f - e^{tA^{(m)}} f|| \to 0 \text{ as } n, m \to \infty, \ \forall f \in D_A. \tag{A.6}$$

Now define the operators T_t by

$$T_t f = \lim_{n \to \infty} e^{tA^{(n)}} f. \qquad (f \in D_A)$$

Since $||T_t f|| \leq \overline{\lim}_{n \to \infty} ||e^{tA^{(n)}} f|| \leq ||f||$, $||T_t|| \leq 1$. Hence T_t can be uniquely extended to $\mathscr{X}$ as a contraction. Moreover,

$$T_{t+s} f = \lim_{n \to \infty} e^{(t+s)A^{(n)}} f = \lim_{n \to \infty} e^{tA^{(n)}} \cdot e^{sA^{(n)}} f = T_t T_s f.$$

Now for $f \in D_A$, (see (2.45) applied to the semigroup $e^{tA^{(n)}}$)

$$\begin{aligned} T_t f - f &= \lim_{n \to \infty} e^{tA^{(n)}} f - f \\ &= \lim_{n \to \infty} \int_0^t e^{sA^{(n)}} A^{(n)} f ds \\ &= \int_0^t T_s Af ds \qquad \text{(since } ||e^{sA^{(n)}}|| \leq 1, \text{ and } A^{(n)} f \to f). \end{aligned} \tag{A.7}$$

Hence $t \to T_t f$ is continuous on $[0, \infty)$ $\forall f \in D_A$. Therefore, $\lim_{t \downarrow 0} T_t f = f$ $\forall f \in D_A$. Since $\overline{D_A} = \mathscr{X}$, it follows that $\mathscr{X}_0 = \mathscr{X}$. Also, (A.7) implies that the domain of the infinitesmal generator B, say, of $\{T_t : 0 \leq t < \infty\}$ contains D_A and

on D_A one has $B = A$. Since $1 \in \rho(A)$, $1 \in \rho(B)$ (by the *necessity* part), $I - B$, and $I - A$ are *one-to-one* on their respective domains, *agree on* D_A, and are *onto* $\mathscr{X}$. Hence $D_A = D_B$. ■

Lemma 1 *Let A and B be bounded operators which commute. If $||e^{tA}|| \leq 1$, $||e^{tB}|| \leq 1$ $\forall t \geq 0$, then $\forall t \geq 0$,*

$$||e^{tA}f - e^{tB}f|| \leq t||Af - Bf|| \qquad\qquad \forall f \in \mathscr{X}.$$

Proof For every positive integer n, one may write

$$e^{tA}f - e^{tB}f = \sum_{k=1}^{n} e^{\frac{k-1}{n}tA} e^{\frac{n-k}{n}tB} \left(e^{\frac{t}{n}A}f - e^{\frac{t}{n}B}f \right).$$

Hence

$$||e^{tA}f - e^{tB}f|| \leq n||e^{\frac{t}{n}A}f - e^{\frac{t}{n}B}f||. \tag{A.8}$$

But

$$||\frac{e^{\frac{t}{n}A} - I}{t/n} - A|| \to 0, \qquad\qquad ||\frac{e^{\frac{t}{n}B} - I}{t/n} - B|| \to 0.$$

Hence

$$n||e^{\frac{t}{n}A}f - e^{\frac{t}{n}B}f|| = t||\frac{e^{\frac{t}{n}A}f - f}{t/n} - \frac{e^{\frac{t}{n}B}f - f}{t/n}|| \to t||Af - Bf||.$$

Therefore, taking the limit in (A.8) as $n \to \infty$, the desired inequality is obtained. ■

The notion of a closed operator given in Definition 2.2 is particularly useful in the context of linear operators on a Banach space. Its proof can be based on the Baire category theorem for metric spaces. Along the way we also prove the open mapping theorem, a useful tool for checking continuity of inverse maps to continuous linear bijections. Let us start with,

Theorem A.2 (Baire Category Theorem) *Let $\mathscr{X}$ be a complete metric space with metric ρ. If $U_1, U_2, \ldots$ are dense open subsets of $\mathscr{X}$, then their intersection $\cap_{n=1}^{\infty} U_n$ is dense in $\mathscr{X}$. In particular, $\cap_{n=1}^{\infty} U_n \neq \emptyset$.*

Proof We wish to show that $\cap_{n=1}^{\infty} U_n \neq \emptyset$ meets any non-empty open $V \subset \mathscr{X}$. By hypothesis, there is a $u_1 \in U_1 \cap V$ and r_1 chosen sufficiently small such that $\overline{B(u_1 : r_1)} \subset U_1 \cap V$, $0 < r_1 < 1$. Proceed inductively for $n \geq 2$ to get $u_n \in U_n$, $0 < r_n < 1/n$, such that $\overline{B(u_n : r_n)} \subset U_n \cap B(u_{n-1} : r_{n-1})$. Then $\{u_n\}$ is a Cauchy sequence since $\rho(u_k, u_m) < 2r_n < 2/n$ for $k, m > n$. By completeness of $\mathscr{X}$, there

is a $u \in \mathscr{X}$ such that $u_n \to u$. For every $m > n$, $u_m \in B(u_n : r_n) \subset \overline{B(u_n : r_n)} \subset U_n$. So $u \in U_n$ for all n. Moreover, $u \in \overline{B(u_1 : r_1)}$ and hence $u \in V$. ■

The proof of the next theorem is aided by a lemma.

Lemma 2 *Let $A : \mathscr{X} \to \mathscr{Y}$ be a bounded, linear, surjective map between Banach spaces $\mathscr{X}$, $\mathscr{Y}$. Then there is an $r > 0$ such that*

$$B_{\mathscr{Y}}(0 : r) \subset A(B_{\mathscr{X}}(0 : 1)),$$

where the subscript label $\ell = \mathscr{X}$ or $\mathscr{Y}$ denotes the space in which the indicated open ball $B_\ell(0 : t)$ is defined, $t = r, 1$.

Proof Since A is linear and surjective, one may express $\mathscr{Y} = \cup_{n=1}^{\infty} \overline{A(B_{\mathscr{X}}(0 : n))}$ as a union of closed sets. It follows from the Baire category theorem that $\overline{A(B_{\mathscr{X}}(0 : n))}$ has a non-empty interior for some n. Thus, scaling by $1/n$, the interior of $\overline{A(B_{\mathscr{X}}(0 : 1))}$ is non-empty. Thus, there is a $y_0 \in \mathscr{Y}$ and $r > 0$ such that

$$y_0 \in B_{\mathscr{Y}}(y_0 : 4r) \subset \overline{A(B_{\mathscr{X}}(0 : 1))}.$$

By symmetry one also has $-y_0 \in \overline{A(B_{\mathscr{X}}(0 : 1))}$. Noting $B_{\mathscr{Y}}(y_0 : 4r) = y_0 + B_{\mathscr{Y}}(0 : 4r)$ and since each $x \in B_{\mathscr{Y}}(0; 4r)$ may be expressed $x = (y_0 + x) + (-y_0)$, it follows that

$$B_{\mathscr{Y}}(0; 4r) \subset \overline{A(B_{\mathscr{X}}(0 : 1))} + \overline{A(B_{\mathscr{X}}(0 : 1))} = 2\overline{A(B_{\mathscr{X}}(0 : 1))},$$

with $R + S := \{r + s : r \in R, s \in S\}$, $R, S \subset \mathscr{Y}$, and the last equality derived from convexity, i.e., for convex C, $C + C = 2C$ (Exercise). Therefore, again by scaling, one has

$$B_{\mathscr{Y}}(0 : 2r) \subset \overline{A(B_{\mathscr{X}}(0 : 1))}. \tag{A.9}$$

Now, let $y \in B_{\mathscr{Y}}(0 : r)$. We seek $x \in B_{\mathscr{X}}(0 : 1)$ such that $Ax = y$. Using, (A.9) for $2y \in B_{\mathscr{Y}}(0 : 2r)$ and $\varepsilon > 0$, there is a $z \in \mathscr{X}$ such that $||2z||_{\mathscr{X}} < 1$, and $||2y - A2z||_{\mathscr{Y}} < 2\varepsilon$. So, let $z_1 \in \mathscr{X}$ such that $||z_1||_{\mathscr{X}} < 1/2$, and $||Az_1 - y||_{\mathscr{Y}} < r/2$. Now iterate application of (A.9) for $4(Az_1 - y) \in B_{\mathscr{Y}}(0 : 2r)$ to obtain $z_2 \in \mathscr{X}$ such that $||z_2||_{\mathscr{X}} < 1/4$, $||A(z_1 + z_2) - y||_{\mathscr{Y}} < r/4$, and so on. This procedure yields $\{z_n\}$ in $\mathscr{X}$ such that $||z_n||_{\mathscr{X}} < 2^{-n}$, $||A(z_1 + \cdots + z_n) - y||_{\mathscr{Y}} < r2^{-n}$. In particular, $\{z_1 + \cdots + z_n\}$ being Cauchy in the Banach space $\mathscr{X}$, there is a $z \in \mathscr{X}$ such that $\lim_n(z_1 + \cdots + z_n) = z \in B(0 : 1)$, and $Az = y$. ■

Theorem A.3 (Open Mapping Theorem) *If $A : \mathscr{X} \to \mathscr{Y}$ is a bounded linear, surjective map between two Banach spaces $\mathscr{X}$, $\mathscr{Y}$, then A is an open map, i.e., A maps open subsets of $\mathscr{X}$ onto open subsets of $\mathscr{Y}$.*

Proof Let $U \subset \mathscr{X}$ be an open set and $y \in A(U)$, say $y = Ax$, $x \in U$. Let us show that there is an open ball containing y and contained in $A(U)$. Choose $r > 0$ such that $x + B_{\mathscr{X}}(0 : r) \subset U$, and therefore $y + A(B_{\mathscr{X}}(0 : r)) \subset A(U)$. By Lemma 2, there is a $t > 0$ such that $B_{\mathscr{Y}}(0 : t) \subset A(B_{\mathscr{X}}(0 : r))$. Thus $A(U) \supset y + B_{\mathscr{Y}}(0 : t) = B_{\mathscr{Y}}(y : t)$, i.e., $A(U)$ is open. ∎

Note that if $A : \mathscr{X} \to \mathscr{Y}$ is a bounded linear map on normed linear spaces then, by continuity, the graph $\{(x, Ax) : x \in \mathscr{X}\}$ is a closed subset of the product space $\mathscr{X} \times \mathscr{Y}$ given the product topology (Exercise). In the case of Banach spaces, the converse holds, and, among other things, this provides an alternative approach to proving that a linear map is bounded.

Theorem A.4 (Closed Graph Theorem) *Let A be a linear map on a Banach space $\mathscr{X}$ into a Banach space $\mathscr{Y}$. If the graph $G_A = \{(x, Ax) : x \in \mathscr{X}\}$ is a closed subset of the product space $\mathscr{X} \times \mathscr{Y}$, then A is a bounded linear map.*

Proof Since G_A is a closed subspace of the Banach space $\mathscr{X} \times \mathscr{Y}$ for the norm $|||(x, y)||| = ||x||_{\mathscr{X}} + ||y||_{\mathscr{Y}}$, G_A is also a Banach space for the norm $||| \cdot |||$. The projection of G_A onto $\mathscr{X}$ given by $\hat{A}(x, Ax) = x$, $x \in \mathscr{X}$ defines a bounded, linear bijection, i.e., one-to-one and onto. Thus, $\hat{A}$ is open by Theorem A.3, and hence $\hat{A}$ is continuous. Therefore, its second coordinate projection A is also continuous. ∎

References

Aldous D, Shields P (1988) A diffusion limit for a class of randomly-growing binary trees. Prob Thry Rel Fields 79(4):509–542

Appuhamillage T, Sheldon D (2012) First passage time of skew Brownian motion. J Appl Probab 49(3):685–696

Appuhamillage TA, Bokil VA, Thomann E, Waymire E, Wood BD (2010) Solute transport across an interface: a Fickian theory for skewness in breakthrough curves. Water Resour Res 46(7):W07511. https://doi.org//10.1029/2009WR008258

Appuhamillage T, Bokil V, Thomann E, Waymire E, Wood B (2011) Occupation and local times for skew Brownian motion with applications to dispersion across an interface. Ann Appl Probab 21(1):183–214

Appuhamillage TA, Bokil VA, Thomann EA, Waymire EC, Wood BD (2014) Skew disperson and continuity of local time. J Stat Phys 156(2):384–394

Aris R (1956) On the dispersion of a solute in a fluid flowing through a tube. Proc Roy Soc Ser A 235:67–77

Arnold L (1974) Stochastic differential equations: theory and applications. Wiley-Interscience, New York

Athreya KB (1985) Discounted branching random walks. Adv Appl Probab 17(1):53–66

Athreya KB, Ney P (1978) A new approach to the limit theory of recurrent Markov chains. Trans Amer Math Soc 245:493–501

Avellaneda M, Majda AJ (1992) Mathematical models with exact renormalization for turbulent transport. II. Fractal interfaces, non-Gaussian statistics and the sweeping effect. Comm Math Phys 146:139–204

Azais R, Bouguet F (Eds) (2018) Statistical inference for piecewise-deterministic Markov processes. John Wiley & Sons

Bacaer N (2011) Verlulst and the logistic equation (1838). In: A short history of mathematical population dynamics. Springer, London

Barndorff-Nielsen O, Halgreen C (1977) Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions. Zeit Wahr und verw Gebiete 38(4):309–311

Barndorff-Nielsen O, Blaesild P, Halgreen C (1978) First hitting time models for the generalized inverse Gaussian distribution. Stochast Proc Appl 7(1):49–54

Barndorff-Nielsen OE, Mikosch T, Resnick SI (Eds) (2001) Lévy processes: theory and applications. Springer Science & Business Media, Berlin

Basak GK, Bhattacharya RN (1992) Stability in distribution for a class of singular diffusions. Ann Probab 20(1):312–321

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8

Bass R (2014) A stochastic differential equation with a sticky point. Electron J Probab 19(32):1–22
Batchelor GK (1952) The effect of homogeneous turbulence on material lines and surfaces. Proc R Soc Lond Ser A Math Phys Sci 213(1114):349–366
Bell J (2015) The Dunford-Pettis theorem, University of Toronto
Ben Arous G, Owhadi H (2003) Multiscale homogenization with bounded ratios and anomalous slow diffusion. Comm Pure Appl. Math. A J Issue Courant Instit Math Sci 56(1):80–113
Bensoussan A, Lions JL, Papanicolaou G (1978) Asymptotic analysis for periodic structures, vol 374. American Mathematical Society, Providence
Berkowitz B, Cortis A, Dror I, Scher H (2009) Laboratory experiments on dispersive transport across interfaces: the role of flow direction. Water Resour Res 45(2): W02201. https://doi.org//10.1029/2008WR007342
Best K, Pfaffelhuber P (2010) The Aldous-Shields model revisited with application to cellular ageing. Elect Comm Probab 15:475–488
Bhattacharya RN (1978) Criteria for recurrence and existence of invariant measures for multidimensional diffusions. Ann Prob 6:541–553
Bhattacharya R (1985) A central limit theorem for diffusions with periodic coefficients. Ann Prob 13(2):385–396
Bhattacharya R (1999) Multiscale diffusion processes with periodic coefficients and an application to solute transport in porous media. Ann Appl Probab 9:951–1020
Bhattacharya RN (1982) On the functional central limit theorem and the law of the iterated logarithm for Markov processes. Zeit für Wahr und verw Geb 60(2):185–201
Bhattacharya RN, Goetze F (1995) Time scales for Gaussian approximation and its break down under a hierarchy of periodic spatial heterogeneities. Bernoulli 1:81–123. Correction ibid 1996:107–108
Bhattacharya RN, Gupta VK (1984) On the Taylor-Aris theory of solute transport in a capillary. SIAM J Appl Math 44:33–39
Bhattacharya R, Majumdar M (2007) Random dynamical systems: theory and applications. Cambridge University Press, Cambridge
Bhattacharya R, Majumdar M (2022) Extinct or endangered: possible or inevitable. In: Pure and applied functional analysis. Special Issue in Honor of Roy Radner, Zaslavski A, Ali Khan M (eds)
Bhattacharya RN, Ramasubramanian S (1982) Recurrence and ergodicity of diffusions. J Multivar Anal 12(1):95–122
Bhattacharya R, Wasielak A (2012) On the speed of convergence of multidimensional diffusions to equilibrium. Stoch Dyn 12(1):1150003
Bhattacharya R, Waymire E (1990, 2009) Stochastic processes with applications. Wiley, New York; Published (2009) In: Classics in applied mathematics series, vol 61. SIAM, Philadelphia
Bhattacharya R, Waymire E (2016) A basic course in probability theory. Springer, New York. Errata: https://sites.science.oregonstate.edu/~waymire/
Bhattacharya R, Waymire E (2021) Random walk, brownian motion, and martingales. Graduate texts in mathematics. Springer, New York
Bhattacharya R, Waymire E (2022) Stationary processes and discrete parameter Markov processes. Graduate texts in mathematics. Springer, New York
Bhattacharya RN, Gupta VK, Walker H (1989) Asymptotics of solute dispersion in periodic porous media. SIAM J Appl Math 49(1):86–98
Bhattacharya R, Denker M, Goswami A (1999) Speed of convergence to equilibrium and to normality for diffusions with multiple periodic scales. Stoch Proc Appl 80(1):55–86
Bhattacharya R, Thomann E, Waymire E (2001) A note on the distribution of integrals of geometric Brownian motion. Stat Probab Lett 55(2):187–192
Billingsley P (1961) The Lindeberg-Levy theorem for martingales. Proc Amer Math Soc 12(5):788–792
Billingsley P (1968) Convergence of probability measures. Wiley, New York
Billingsley P (1995) Probability and measure, 3rd edn. Wiley, New York

Blackwell D (1958) Another countable Markov process with only instantaneous states. Ann Math Statist 29(1):313–316

Bokil VA, Gibson NL, Nguyen SL, Thomann EA, Waymire EC (2020) An Euler-Maruyama method for diffusion equations with discontinuous coefficients and a family of interface conditions. J Comp App Math 368:112545

Bose A, Dasgupta A, Rubin H (2002) A contemporary review and bibliography of infinitely divisible distributions and processes. Sankhyā Indian J Statist Ser A 64:763–819

Breiman L (1968) Probability. Addison Wesley, Reading. SIAM, Philadelphia

Brooks JK, Chacon RV (1983) Diffusions as a limit of stretched Brownian motions. Adv Math 49(2):109–122

Cameron RH, Martin WT (1944) An expression for the solution of a general class of non-linear integral equations. Amer J Math 66:281–298

Cameron RH, Martin WT (1945) Transformations of Wiener integrals under a general class of linear transformations. Trans Amer Math Soc 58:184–219

Chen MF, Wang FY (1995) Estimation of the first eigenvalue of second order elliptic operators. J Funct Anal 131(2):345–363

Chen MF (2006) Eigenvalues, inequalities, and ergodic theory. Springer Science & Business Media, Berlin

Chen L, Dobson S, Guenther R, Orum C, Ossiander M, Thomann E, Waymire E (2003) On Itô's complex measure condition. Institute of Mathematical Statistics lecture notes-monograph series, pp 65–80

Cherny AS, Shiryaev AN, Yor M (2002) Limit behavior of the "horizontal–vertical"random walk and some extensions of the Donsker–Prokhorov invariance principle.Th Probab Appl 47(3):498–516

Chung KL (1967) Markov chains with stationary transition probabilities. Springer

Comtet A, Monthus C, Yor M (1998) Exponential functionals of Brownian motion and disordered systems. J Appl Probab 35:255–271

Costa OLV (1990) Stationary distributions for piecewise-deterministic Markov processes. J Appl Probab 27(1):60–73

Crump KS (1975) On point processes having an order statistic structure. Sankhya Indian J Statist Ser A 37:396–404

Crudu A, Debussche A, Muller A, Radulescu O (2012) Convergence of stochastic gene networks to hybrid piecewise deterministic processes. Ann Appl Probab 1822–1859

Csáki E, Hu Y (2003) Lengths and heights of random walk excursions. Discrete random walks, DRW03, Paris, France, pp 45–52. hal-01183932

Dascaliuc R, Michalowski N, Thomann E, Waymire E (2018a) A delayed Yule process. Proc Amer Math Soc 146(3):1335–1346

Dascaliuc R, Thomann EA, Waymire EC (2018b) Stochastic explosion and non-uniqueness for α-Riccati equation. J Math Anal Appl 476(1):53–85. Errata (2023): https://arxiv.org/abs/2303.05482

Dascaliuc R, Pham TN, Thomann E, Waymire EC (2023a) Doubly stochastic Yule cascades (part I): the explosion problem in the time-reversible case. J Funct Anal 284(1):109722

Dascaliuc R, Pham TN, Thomann E, Waymire EC (2023b) Doubly stochastic Yule cascades (part II): the explosion problem in the non-reversible case. Ann de'l Inst Henri Poincaré (to appear)

Dascaliuc R, Pham T, Thomann EA, Waymire EC (2023c) Errata to 'Stochastic explosion and non-uniqueness for α-Riccati equation'. J Math Anal Appl 476(1). https://arxiv.org/abs/2303.05482

Davies PL (1986) Rates of convergence to the stationary distribution for k-dimensional diffusion processes. J Appl Prob 23(2):370–384

Davis MHA (1984) Piecewise-deterministic Markov processes: a general class of non-diffusion stochastic models. J R Statist Soc Ser B (Methodol) 46(3):353–388

Davis M, Etheridge A (2006) Louis Bachelier's theory of speculation: the origins of modern finance. Princeton University Press, Princeton

Diaconis P, Stroock D (1991) Geometric bounds for eigenvalues of Markov chains. Ann Appl Probab 1:36–61

Diaconis P, Freedman D (1999) Iterated random functions. SIAM Rev 41(1):45–76

Dobrushin RL (1956) An example of a countable homogeneous Markov process all states of which are instantaneous (Russian). Teor Veroyatnost i Primenen 1:481–485

Dong C, Iksanov O, Pilipenko A (2023) On a discrete approximation of a skew stable Lé vy process. arXiv:2302.07298

Durrett R (1996) Stochastic calculus: a practical introduction, vol 6. CRC Press, Boca Raton

Durrett R, Kesten H, Waymire E (1991) On weighted heights of random trees. J Theor Probab 4(1):223–237

Dunford N, Schwartz JT (1963, 1988) Linear operators, part 1: general theory, vol 10. Wiley, New York

Dynkin EB (1965) Markov processes, vol 1. Springer, Berlin

Einstein A (1905) On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat (English translation, 1956). Investigations on the theory of the Brownian movement. Ann Phys 322:549560

Ethier SN, Kurtz TG (1985) Markov processes: characterization and convergence. Wiley, New York

Feller W (1940) On the integro-differential equations of purely discontinuous Markoff processes. Trans Amer Math Soc 48(3):488–515

Feller W (1952) The parabolic differential equations and the associated semigroups of transformations. Ann Math 55:468–519

Feller W (1954) Diffusion processes in one dimension. Trans Amer Math Soc 77(1):1–31

Feller W (1957) Generalized second order differential operators and their lateral conditions. Ill J Math 1(4):459–504

Feller W (1968, 1971) An introduction to probability theory and its applications, vol 1, 3rd edn, vol 2, 2nd edn. Wiley, New York

Feller W, McKean HP (1956) A diffusion equivalent to a countable Markov chain. Proc Nat Acad Sci 42(6):351–354

Fill JA (1991) Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. Ann Appl Probab 1:62–87

Folland GB (1984) Real analysis: modern techniques and their applications, vol 40. Wiley, Hoboken

Franke B (2004) Integral inequalities for the fundamental solutions of diffusions on manifolds with divergence-free drift. Mathematische Zeitschrift 246(1):373–403

Freedman D (1971) Brownian motion and diffusion. Holden-Day, New York. Reprint Springer, New York

Freedman D (1983) Constructing the general Markov chain. In: Approximating countable Markov chains. Springer, New York

Fried JJ, Combarnous MA (1971) Dispersion in porous media. In: Advances in hydroscience, vol 7, pp 169–282. Elsevier, Amsterdam

Friedman A (1964) Partial differential equations of parabolic type. Reprinted (2008), Courier Dover Publications, Mineola

Friedman A (1973) Wandering out to infinity of diffusion processes. Trans Amer Math Soc 184:185–203

Friedman A (1975) Stochastic differential equations and applications. Academic, New York

Gelhar LW, Axness CL (1983) Three-dimensional stochastic analysis of macrodispersion in aquifers. Water Resour Res 19(1):161–180

Gilpin ME, Ayala FJ (1973) Global models of growth and competition. Proc Natl Acad Sci USA 70(12):Part I, 3590–3593

Girsanov IV (1960) On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. Thry Probab Its Appl 5(3):285–301

Gompertz B (1825) XXIV. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. FRS & c. Philos Trans R Soc Lond 115:513–583

Goncalves B, Huillet T, Löcherbach E (2022) On population growth with catastrophes. Stoch Models 38(2):214–249

Gupta VK, Mesa OJ, Waymire E (1990) Tree-dependent extreme values: the exponential case. J Appl Probab 27(1):124–133

Handwerk A, Willems H (2007) Wolfgang Doeblin: a mathematician rediscovered. Springer, Berlin

Hanson FB, Tuckwell HC (1978) Persistence times of populations with large random fluctuations. Theoret Populat Biol 14:46–61

Hanson FB, Tuckwell HC (1981) Logistic growth with random density independent disasters. Theoret Populat Biol 19:1–18

Hanson FB, Tuckwell HC (1997) Population growth with randomly distributed jumps. J Math Biol 36:169–187

Harrison M, Shepp L (1981) On skew Brownian motion. Ann Probab 9(2):309–313

Henshaw K (2022) Mathematical perspectives on insurance for low-income populations. Doctoral Dissertation, University of Liverpool, UK

Hoteit H, Mose R, Younes A, Lehmann F, Ackerer P (2002) Three-dimensional modeling of mass transfer in porous media using the mixed hybrid finite elements and the random-walk methods. Math Geol 34(4):435–456

Ibragimov IA (1963) A central limit theorem for a class of dependent random variables. Thry Probab Appl 8(1):83–89

Ikeda N, Watanabe S (1989) Stochastic differential equations and diffusion processes, 2nd edn. North-Holland, Amsterdam

Iksanov A, Pilipenko A (2023) On a skew stable Lévy process. Stochast Proc Appl 156:44–68

Itô K (2004) Stochastic processes: lectures given at Aarhus University. Springer Science & Business Media, Berlin

Itô K (1965) Generalized uniform complex measures in the Hilbertian metric space with the application to the Feynman integral. Proc Fifth Berkeley Symp Math Stat Probab II:145–161

Itô K, McKean HP (1963) Brownian motions on a half line, Illinois. J Math 7:181–231

Itô K, McKean HP (1965) Diffusion processes and their sample paths. Reprint Springer, Berlin

Itô K, Rao KM (1961) Lectures on stochastic processes, vol 24. Tata Institute of Fundamental Research, Bombay

Karatzas I, Shreve SE (1991) Brownian motion and stochastic calculus, 2nd edn. Springer, New York

Karlin S, Taylor HE (1981) A second course in stochastic processes. Elsevier, Amsterdam

Khaśminskii RZ (1960) Ergodic properties of recurrent diffusions and stabilization of the Cauchy problem for parabolic equations, Teoriya Veroyat Primen 5:7–28

Kingsland S (1982) The refractory model: the logistic curve and the history of population ecology. Quart Rev Biol 57(1):29–52

Lande R, Engen S, Saether B-E (2003) Stochastic population dynamics in ecology and conservation, Oxford series in ecology and evolution. Oxford University Press, New York, p 396

Lawler GF (2008) Conformally invariant processes in the plane, no 114, American Mathematical Society, Providence

Le Jan YL, Sznitman AS (1997) Stochastic cascades and 3-dimensional Navier-Stokes equations. Probab Thry Rel Fields 109(3):343–366

Lejay A (2006) On the constructions of the skew Brownian motion. Probab Surv 3:413–466

Lejay A, Martinez M (2006) A scheme for simulating one-dimensional diffusion processes with discontinuous coefficients. Ann Appl Probab 16(1):107–139

Lévy P (1948) Processus stochastiques et mouvement brownien. Gauthier-Villars, Paris

Lévy P (1951) Systems Markoviens et stationnaires, Cas denombrable. Ann Sci de l'Ecole Normale Superieure de Paris 68:327–381

Lévy P (1953) Random functions: general theory with special reference to Laplacian random functions, vol 1, no 12. University of California Press, Chicago
Lévy P (1954) Théorie de l'addition des variables aléatories, 2nd edn. Gauthier-Villars, Paris (1st edn. 1937)
Lindvall T (2002) Lectures on the coupling method. Courier Corporation, North Chelmsford
Majda AJ, Tong XT (2016) Geometric ergodicity for piecewise contracting processes with applications for tropical stochastic lattice models. Commun Pure Appl Math 69(6):1110–1153
Mandl P (1968) Analytical treatment of one-dimensional Markov processes, vol 151. Springer, Berlin
Mandelbrot BB, Van Ness JW (1968) Fractional Brownian motions, fractional noises and applications. SIAM Rev 10(4):422–437
Mao YH (2006) Convergence rates in strong ergodicity for Markov processes. Stoch Proc Appl 116(12):1964–1976
McKean HP (1969) Stochastic integrals. Academic, New York
Metzler A (2013) The Laplace transform of hitting times of integrated geometric Brownian motion. J Appl Probab 50(1):295–299
Meyer PA (1966) Probability and potentials, vol 1318. Blaisdell Publishing Company, Waltham
Meyn S, Tweedie RL (1993) Markov chains and stochastic stability. Cambridge University Press, Cambridge
Novikov A (1972) On an identity for stochastic integrals. Thry Probab Appl 17(4):717–720
Nummelin E (1978) Splitting technique for Harris recurrent Markov chains. Z Wahrs Verw Geb 43:309–318
Obukhov A (1941) Spectral energy distribution in a turbulent flow. Izv Akad Nauk SSSR Ser Geogr i Geofiz 5:453–466
Orey S (1971) Limit theorems for Markov chain transition probabilities. van Nostrand, London
Ouknine Y (1991) Skew-Brownian motion" and derived processes. Thry Prob Its Appl 35(1):163–169
Owhadi H (2003) Anomalous slow diffusion from perpetual homogenization. Ann Prob 31(4):1935–1969
Peckham SD, Waymire EC, De Leenheer P (2018) Critical thresholds for eventual extinction in randomly disturbed population growth models. J Math Bio 77(2):495–525
Peskir G (2015) William feller, selected papers II. Springer, Berlin, pp 77–93
Pinsky RG, Dieck TT (1995) Positive harmonic functions and diffusion, vol 45. Cambridge University Press, Cambridge
Pitman J, Yor M (2001) On the distribution of ranked heights of excursions of a Brownian bridge. Ann Probab 29(1):361–384
Ramirez JM (2012) Population persistence under advection-diffusion in river networks. J Math Bio 65(5):919–942
Ramirez JM, Thomann EA, Waymire EC, Haggerty R, Wood B (2006) A generalized Taylor-Aris formula and skew diffusion. Multiscale Model Simulat 5(3):786–801
Ramirez JE, Thomann EA, Waymire EC (2013) Advection-disperson across interfaces. Stat Surv 28(4):487–509
Ramirez JM, Thomann EA, Waymire EC (2016) Continuity of local time: an applied perspective. In: The fascination of probability, statistics and their applications; In: Honour of Ole E. Barndorff-Nielsen, Podolskij M, Steltzer R, Thorbjornsen S, Veraart AED (eds). Springer, Berlin
Rao KM (1969) On decomposition theorems of Meyer. Math Scandinavica 24(1):66–78
Reed M, Simon B (1972) Methods of modern mathematical physics, vol 1. Academic, New York
Reuter GEH (1957) Denumerable Markov processes and the associated contraction semigroups on l. Acta Mathematica 97:1–46
Revuz D, Yor M (1999) Continuous martingales and brownian motion, 3rd edn. Springer, Berlin
Richards FJ (1959) A flexible growth function for empirical use. J Exper Botany 10:290–300
Rogers LC, Pitman JW (1981) Markov functions. Ann Probab 9:573–582
Rogers LCG, Shi Z (1995) The value of an Asian option. J Appl Probab 32(4):1077–1088

Salminen P, Stenlund D (2021) On occupation times of one-dimensional diffusions. J Theoret Probab 34(2):975–1011

Scheutzow M (2019) Lecture notes in stochastic processes III, BMS advanced course. http://page.math.tu-berlin.de/~scheutzow/WT4main.pdf

Schlomann BH (2018) Stationary moments, diffusion limits, and extinction times for logistic growth with random catastrophes. J Theor Bio 454:154–163

Schreiber SJ (2012) Persistence for stochastic difference equations: a mini-review. J Differ Equ Appl 18(8):1381–1403

Skorokhod AV (1961) Stochastic equations for diffusion processes in a bounded region, I. Theory Probab Appl 6(3):264–274

Skorokhod AV (1962) Stochastic equations for diffusion processes in a bounded region, II. Theory Probab Appl 7(1):3–23

Stein EM (1970) Singular integrals and differentiability properties of functions. Princeton University Press, Princeton

Stroock DW, Varadhan SS (1969) Diffusion processes with continuous coefficients, I; II. Comm Pure Appl Math 22(3):345–400; 22(4):479–530

Stroock DW, Varadhan SRS (1979) Multidimensional diffusion processes. Springer, Berlin

Tanaka H (1963) Note on continuous additive functionals of the 1-dimensional Brownian motion Path. Z Wahr verw Gebiete 1:251–257

Taylor GI (1953) Dispersion of Soluble matter in a solvent flowing through a tube. Proc R Soc Ser A 219:186–203

Thomann E, Waymire EC (2003) Contingent claims on assets with conversion costs. J Stat Plan Infer 113:403–417

Walsh J (1978) A diffusion with a discontinuous local time. Asterisque 52(53):37–45

Wasielak A (2009) Various limiting criteria for multidimensional diffusion processes, PhD Thesis University of Arizona, Tucson

Winter CL, Neuman SP, Newman CM (1984) Prediction of far-field subsurface radionuclide dispersion coefficients from hydraulic conductivity measurements: a multidimensional stochastic theory with application to fractured rocks (No. NUREG/CR-3612), Arizona University, Tucson (USA), Department of Hydrology and Water Resources

Wooding RA (1960) Rayleigh instability of a thermal boundary layer in flow through a porous medium. J Fluid Mech 9:183–192

Woyczyński WA (2001) Lévy processes in the physical sciences. In: Lévy processes. Birkhauser, Boston, pp 241–266

Yosida K (1964) Functional analysis, 1st ed. Springer, New York

Related Textbooks and Monographs

The following is list of some supplementary and/or follow-up reading, including the books cited in the text.

Arnold L (1974) Stochastic differential equations: theory and applications. Wiley-Interscience, New York

Asmussen S, Hering H (1983) Branching processes. Birkhäuser, Boston

Athreya KB, Ney PE (1972) Branching processes. Springer, New York

Bensoussan A, Lions JL, Papanicolaou G (1978) Asymptotic analysis for periodic structures, vol 374. American Mathematical Society, Providence

Bhattacharya R, Waymire E (1990) Stochastic processes with applications. Wiley, New York; Published (2009) In: Classics in applied mathematics series, vol 61, SIAM, Philadelphia

Bhattacharya R, Majumdar M (2007) Random dynamical systems: theory and applications. Cambridge University Press, Cambridge

Bhattacharya R, Waymire E (2016) A basic course in probability. Springer, New York. Errata: https://sites.science.oregonstate.edu/~waymire/

Bhattacharya R, Waymire E (2021) Random walk, brownian motion, and martingales. Graduate texts in mathematics. Springer, New York

Bhattacharya R, Waymire E (2022) Stationary processes and discrete parameter Markov processes. Graduate texts in mathematics. Springer, New York

Billingsley P (1968) Convergence of probability measures. Wiley, New York

Billingsley P (1995) Probability and measure, 3rd edn. Wiley, New York

Breiman L (1968) Probability. Addison Wesley, Reading. Reprint SIAM, Philadelphia

Chen MF (2006) Eigenvalues, inequalities, and ergodic theory. Springer Science & Business Media, Berlin

Chung KL (1967) Markov chains with stationary transition probabilities. Springer

Chung KL, Williams RJ (1990) Introduction to stochastic integration, 2nd edn. Birkhauser, Boston

Dieudonné J (1960) Foundations of modern analysis. Academic, New York

Doob JL (1953) Stochastic processes. Wiley, New York

Dunford N, Schwartz JT (1963, 1988) Linear operators, part 1: general theory, vol 10. Wiley, New York

Durrett R (1984) Brownian motion and martingales in analysis. Wadsworth, Belmont

Durrett R (1995) Probability theory and examples, 2nd edn. Wadsworth, Brooks & Cole, Pacific, Grove

Durrett R (1996) Stochastic calculus: a practical introduction, vol 6. CRC Press, Boca Raton

Dym H, McKean HP (1972) Fourier series and integrals. Academic, New York

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8

Dynkin EB (1965) Markov processes, vol 1, 2. Springer, Berlin
Ethier SN, Kurtz TG (1985) Markov processes: characterization and convergence. Wiley, New York
Feller W (1968, 1971) An introduction to probability theory and its applications, vol 1. 3rd edn, vol 2, 2nd edn. Wiley, New York
Folland GB (1984) Real analysis: modern techniques and their applications, vol 40. Wiley, Hoboken
Freedman D (1971) Brownian motion and diffusion. Holden-Day, New York. Reprint Springer, Berlin
Freedman D (1983) Constructing the general Markov chain. In: Approximating countable Markov chains. Springer, New York
Friedman A (1964) Partial differential equations of parabolic type. Reprinted Courier Dover Publications, Mineola (2008)
Friedman A (1969) Partial differential equations, vol 1. Holt, Reinhart and Winston, New York, p 969
Friedman A (1975) Stochastic differential equations and applications. Academic, New York
Gihman II, Skokohod AV (1974) The theory of stochastic processes I. Springer, New York. English translation by S. Kotz of the original Russian published in 1971 by Nauka, Moscow
Gilbarg D, Trudinger NS (1977) Elliptic partial differential equations of second order, vol 224, no 2. Springer, Berlin
Hall P, Heyde CC (1980) Martingale limit theory and its application. Academic, New York
Harris TE (1963) The theory of branching processes. Springer, Berlin
Ikeda N, Watanabe S (1989) Stochastic differential equations and diffusion processes, 2nd edn. North-Holland, Amsterdam
Itô K (2004) Stochastic processes: lectures given at Aarhus University Springer Science & Business Media, Berlin
Itô K, McKean HP (1965) Diffusion processes and their sample paths. Reprint Springer, Berlin
Itô K, Rao KM (1961) Lectures on stochastic processes, vol 24. Tata Institute of Fundamental Research, Bombay
Jagers P (1975) Branching processes with applications to biology. Wiley, New York
Kallenberg O (2001) Foundations of modern probability, 2nd edn. Springer, New York
Karatzas I, Shreve SE (1991) Brownian motion and stochastic calculus, 2nd edn. Springer, New York
Karlin S, Taylor HM (1975) A first course in stochastic processes. Academic, New Yok
Karlin S, Taylor HM (1981) A second course in stochastic processes. Elsevier, Amsterdam
Khaśminskii R (2011) Stochastic stability of differential equations, vol 66. Springer Science& Business Media, Berlin
Lawler G (2008) Conformally invariant processes in the plane. American Mathematical Society, Providence
Lévy P (1925) Calcul des probabilités. Gauthier-Villars, Paris
Lévy P (1948) Processus stochastiques et mouvement brownien. Gauthier-Villars, Paris
Lévy P (1954) Théorie de l'addition des variables aléatories, 2nd edn. Gauthier-Villars, Paris. (1st edn. (1937))
Liggett TM (1985) Interacting particle systems. Springer, New York
Liggett TM (2010) Continuous time markov processes: an introduction. Graduate studies in mathematics, vol 112. American Mathematical Society, Providence
Lindvall T (2002) Lectures on the coupling method. Courier Corporation, North Chelmsford
Mandl P (1968) Analytical treatment of one-dimensional Markov processes, vol 151. Springer, Berlin
McKean HP (1969) Stochastic integrals. Academic, New York
Meyer PA (1966) Probability and potentials, vol 1318. Blaisdell Publishing Company, Waltham
Meyn S, Tweedie RL (1993) Markov chains and stochastic stability. Cambridge University Press, Cambridge

Neveu J (1971) Mathematical foundations of the calculus of probability. Holden-Day, San Francisco

Neveu J (1975) Discrete parameter martingales. North-Holland, Amsterdam

Nualart D (1995) The malliavin calculus and related topics. Springer, New York

Oksendal B (1998) Stochastic differential equations, 5th edn. Springer, Berlin

Orey S (1971) Limit theorems for Markov chain transition probabilities. van Nostrand, London

Parthasarathy KR (1967) Probability measures on metric spaces. Academic, New York

Pinsky RG, Dieck TT (1995) Positive harmonic functions and diffusion, vol 45. Cambridge University Press, Cambridge

Reed M, Simon B (1972) Methods of modern mathematical physics, vol 1. Academic, New York

Revuz D, Yor M (1999) Continuous martingales and brownian motion, 3rd edn. Springer, Berlin

Rogers LCG, Williams D (2000) Diffusions, Markov processes and martingales, vol 1, 2, 2nd edn. Cambridge University Press, Cambridge

Royden HL (1988) Real analysis, 3rd edn. MacMillan, New York

Stein EM (1970) Singular integrals and differentiability properties of functions. Princeton University Press, Princeton

Stroock DW, Varadhan SRS (1979) Multidimensional diffusion processes. Springer, Berlin

Yosida K (1964) Functional analysis, 1st edn. Springer, New York

Symbol Index

Special Sets and Functions:

By convention, $X(t)$ and X_t in continuous parameter, and $X(n)$ and X_n discrete parameter, may be used interchangeably indexing stochastic processes.

In the classic notation of G.H. Hardy, one writes $a(x) = O(b(x))$ to mean that there is a constant c (independent of x) such that $|a(x)| \le c|b(x)|$ for all x. Also $a(x) = o(b(x))$ indicates that the ratio $a(x)/b(x) \to 0$ according to specified limit.

$\mathbb{Z}_+$, set of non-negative integers

$\mathbb{Z}_{++}$, set of positive integers

$\mathbb{R}_+$, set of non-negative real numbers

$\mathbb{R}_{++}$, set of positive real numbers

$\mathbb{R}_0 = \mathbb{R}\backslash\{0\}$

∂A, boundary of set A

A^o, interior of set A

A^-, closure of set A

A^c, complement of set A

$\mathbf{1}_B(x)$, $\mathbf{1}_B$, indicator function of B

$[X \in B]$, inverse image of the set B under X

$\#A$, $|A|$, card(A), cardinality for finite set A

δ_x Dirac delta (point mass)

$\otimes$, σ-field product

$\bigotimes$, product of $\sigma-$fields

$p(t; x, dy)$, homogeneous (stationary) transition probability

$p(s, t; x, dy)$, nonhomogeneous (nonstationary) transition probability

P_x distribution of Markov process with initial value x

$\mathbb{E}_x$ expected value with respect to P_x

$X_t^+ = \{X_t^+(s) \equiv X_{t+s} : s \ge 0\}$, after-$t$ process

$B_s^t = \{B_u - B_s : s \le u \le t\}$

$\mathcal{F}_\tau$ pre-τ σ-field

$\mathcal{B}$, $\mathcal{B}(S)$ Borel σ-field

$\mathcal{B}_0$ collection of Borel subsets of $(0, \infty) \times \mathbb{R}_0$

$\mathcal{R}_0 = \{A \in \mathcal{B}_0 : A \subset (0, \delta) \times \{\Delta : |\Delta| > \delta^{-1}$ for some $\delta > 0\}$

$\mathcal{D}_{st}(X)$ differential σ-field associated with X

$\Delta_{st}X = X(t) - X(s)$

$\rho(A)$, resolvent set for linear operator A

$B(x : r)$, open ball centered at x of radius r; $\overline{B(x : r)}$, the closed ball

$\partial B(x : r)$, the boundary set of $B(x : r)$

Function Spaces, Elements and Operations:

$(R, \mathcal{R})$ generic measurable range space for random map

$\mathbf{K}$ generic path space for stochastic process

$C[0, 1]$, set of continuous, real-valued functions defined on $[0, 1]$

S^∞, infinite product space, space of infinite sequences in S

$C([0, \infty) : S)$, set of continuous functions on $[0, \infty)$ with values in topological space S

$C_b(S)$ or $C_b(S : \mathbb{R})$, set of continuous bounded, real-valued functions on a metric (or topological) space S

$\mathbb{B}(S)$, set of bounded, measurable real-valued functions on a measurable space $(S, \mathcal{S})$

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8

$C_0(S)$ or $C_0(S:\mathbb{R})$ space of continuous real-valued functions on metric space or topological space S that vanish at infinity.
$C(S:\mathbb{C})$, set of continuous, complex-valued functions on S
$D([0,\infty):S)$, $D[0,T]$, Skorokhod spaces of càdl'ag functions
$\mathcal{M}[\alpha,\beta]$, space of real-valued progressively measure stochastic processes $f(t), \alpha \le t \le \beta$, such that $\mathbb{E}\int_\alpha^\beta f^2(t)dt < \infty$
$\mathcal{L}[\alpha,\beta]$, space of real-valued progressively measure stochastic processes $f(t), \alpha \le t \le \beta$, such that $P(\int_\alpha^\beta f^2(s)ds < \infty) = 1$
$\mathcal{N}_A$, the null space (or kernel) of the operator A
$\mathcal{R}_A$, the range space of the operator A
$\langle\cdot,\cdot\rangle$, inner product
$1^\perp = \{f : \langle f, 1\rangle = 0\}$
(gradient vector) $\nabla = \text{grad} = (\frac{\partial}{\partial x_j})_{1\le j\le m}$
(divergence) $\text{div}b(x) = \nabla\cdot b(x) = \sum_{j=1}^m \frac{\partial b_j(x)}{\partial x_j}, b = (b_1,\ldots,b_m)$
(generalized derivative) $D_h f(x) = \frac{df(x)}{dg(x)} = \lim_{h\to 0}\frac{f(x+h)-f(x)}{g(x+h)-g(x)}$, $D_x f(x) = \frac{df(x)}{dx}$.
(Laplacian) $\Delta = \nabla\cdot\nabla = \sum_{j=1}^m \frac{\partial^2}{\partial x_j^2}$
(Hessian matrix) $\nabla'\nabla = ((\frac{\partial^2}{\partial x_i \partial x_j}))_{1\le i,j\le m}$
R_λ, resolvent operator
$\mathcal{X}_0$, center of semigroup on Banach space $\mathcal{X}$
A^*, T_t^*, adjoint operator to A, T_t, respectively
ζ, explosion time
$\ell(t,x)$, local time
e_i **e** i-th coordinate of unit vector **e**
$i.o.$ infinitely often
$f * g$, convolution of functions
$Q_1 * Q_2$, convolution of probabilities
Cov, covariance
Var, variance
$\Pi_{t(1),\ldots,t(m)}(f)$, finite-dimensional projection of f to $(f(t_1),\ldots,f(t_m))$
$\text{sgn}(x) \equiv \frac{x}{|x|} = \begin{cases} 1 & \text{if } x \ge 0 \\ -1 & \text{if } x < 0 \end{cases}$
$\langle M\rangle_T$, quadratic variation of martingale M_t on interval $0 \le t \le T$.
$\Rightarrow$, weak convergence
O_p, little O in probability
'∞' or s_∞, compactification point at infinity
A', v' matrix, including vector, transpose
$\text{tr}(A) = \sum_{i=1}^k a_{ii}$, matrix trace of $A = ((a_{ij}))_{1\le i,j\le k}$
p-lim, limit in probability

Author Index

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8

Subject Index

R. Bhattacharya, E. C. Waymire, *Continuous Parameter Markov Processes and Stochastic Differential Equations*, Graduate Texts in Mathematics 299,
https://doi.org/10.1007/978-3-031-33296-8

S

T